

## Korpora und Belegsammlungen\*

### 0. Vorbemerkungen

0.1. Ein geplantes Unternehmen, das als "neues großes Wörterbuch der deutschen Sprache" gekennzeichnet wird (These 1), ein übernommenes Werk<sup>1</sup> mit der Charakterisierung "neues, groszes wörterbuch der deutschen sprache"<sup>2</sup> weist sich durch das Prädikat "Wörterbuch der deutschen Sprache" bzw. in seinem "streben nach umfassender samlung und behandlung (aller deutschen wörter)"<sup>3</sup> als lexikographisches Unternehmen aus. Als solches gehört es in den Bereich "einer primär-induktiv bzw. empirisch-deskriptiv verfahrenen Linguistik"<sup>4</sup>; es ist "(seinem) ziel und (seiner) bestimmung (nach) . . . im edelsten sinne practisch."<sup>5</sup>

0.2. Die Attribute *neu* und *grosz/groß* drücken aus, daß es sich um "ein unabsehbares, von keinem angelegtes, geschweige vollbrachtes werk" handelt, um "eine weit vollere und lebendigere samlung aller deutschen wörter . . . , als sie noch stattgefunden hat"<sup>6</sup>; ein interdisziplinäres Wörterbuch, in dem die Gemeinsprache und die Fachsprachen "nach dem gleichen linguistischen und lexikographischen Konzept" beschrieben werden, gibt es nicht.<sup>7</sup>

0.3. Einem Großunternehmen wie Grimms Deutschem Wörterbuch werden "wie der geschichte die urkunden, die reichsten sammlungen des sprachvorrats unentbehrlich"<sup>8</sup>; für ein empirisch-deskriptives Großunternehmen wie das geplante Interdisziplinäre deutsche Wörterbuch "ist ein Textkorpus als Datenbasis nicht nur nützlich, sondern unentbehrlich".<sup>9</sup>

0.4. Diesen Parallelen entsprechend geht es im folgenden darum

- die von Grimm erörterten Gesichtspunkte zum Thema 'Quellen für das Deutsche Wörterbuch' kommentierend in eine vom direkten Bezug auf das Deutsche Wörterbuch entbundene Formulierung zu bringen (vgl. 1.)
- diese allgemeineren Gesichtspunkte einzubetten in Überlegungen zum Thema 'Korpora und Belegsammlungen für das Interdisziplinäre deutsche Wörterbuch' (vgl. 2.)

\* Für das 2. Kapitel dieses Aufsatzes ist ein Arbeitspapier von mir mitverwendet worden, das zuerst auf dem 2. Colloquium (Februar 1976) und in 2. Fassung auf dem 4. Colloquium (Juni 1976) in Bad Homburg vorgelegt worden ist. An der 2. Fassung waren H. Bergenholtz, G. Engelen und B. Schader beteiligt, denen ich für kritische und erweiternde Hinweise herzlich danke. Im Oktober 1976 wurde von mir eine 2. erweiterte Fassung fertiggestellt.

## 1. Gesichtspunkte zum Thema 'Quellen für das Deutsche Wörterbuch'

### 1.1. Wörterbuch – Sprache – Literatur

Für Grimm ist bei der Erarbeitung seines Wörterbuchs der Rückgriff auf schriftliche Quellen selbstverständlich.<sup>10</sup> Dies ergibt sich aus der engen Beziehung – aus der Gleichsetzung – von "sprache und literatur", die erst eigentlich 1450 gestiftet wird durch "die grosze erfindung der druckerei", mit der speziell "der sprachwissenschaft ganz neue bahnen gesprengt worden" sind: "Seitdem schriften gedruckt und aller enden gelesen werden, sind wörterbücher entsprungen", die nicht eine "angehäufte wortmasse", sondern "die volle aufstellung der sprachen" darstellen; erst durch "die vervielfältigung der schrift" wird gegenüber früheren Sprachzuständen "die ungleich grözere masse des heutigen sprachschatzes . . . durch überfließen- de belege lehrreich begründet". Deren Berücksichtigung ist Grundvoraussetzung für die genaueste "ermittelung aller einzelnen wörter" und ihre Sammlung, für die Belegung dieser Wörter, für die "auffassung der wortbedeutungen", die breite "entfaltung der bedeutung", und für die Erhellung der Geschichte der Wörter.<sup>11</sup>

Dieser von Grimm so gesehene Zusammenhang zwischen allgemein-verfügbaren literarischen Quellen, Sprache und Wörterbuch spiegelt sich wider in der zusammenfassenden Beschreibung des Objektbereichs, die den Punkt "10. Umfang der Quellen" des Vorworts einleitet: "das wörterbuch (soll) sich über die gesamte hochdeutsche schriftsprache von der mitte des funfzehnten jahrhunderts an bis auf heute . . . erstrecken": Beschränkung auf das Neuhochdeutsche, dabei aber nicht auf "die kurze spanne der gegenwart", sondern mit dem Ziel, "den umfang des nhd. ganzen zeitraums so viel als möglich zu erschöpfen"<sup>12</sup>: Es ist für das neue große Wörterbuch mit<sup>13</sup> konstitutiv, daß auf die Quellen, deren Merkmale mit dem Objektbereich des Wörterbuches beschrieben sind, zurückgegriffen wird und daß diese Quellen in Form von Belegen für die Wortartikel verfügbar gemacht werden: Konstitutive Quellen- und Belegbezogenheit des Wörterbuchs.

### 1.2. Richtlinien und Hilfe für Materialbereitstellung

Die "durchsicht der quellen und anfertigung der auszüge" ist ein großes organisatorisches Problem. Durch die kostendeckende Bereitschaft der Verlagshandlung sind "sehr schätzbare und in der that unentbehrliche samlungen zu stande gekommen"; doch ist zu sagen, daß sie "gleichwohl, ungeachtet dasz ein genauer plan des verfahrens entworfen war . . . , nach beschaffenheit der schriftsteller und nach der ausziehenden anstelligkeit oder neigung von sehr verschiedenem werthe sein musten"<sup>14</sup>: Es ist ratsam, für die Materialbereitstellung klare Richtlinien festzulegen und sich bei der Durch-

führung weiterer Hilfe zu versichern: Notwendige Planung der Materialbereitstellung und Sicherstellung der Hilfeleistungen anderer.

### 1.3. Literarische Quellen

Dem Grundsatz, daß "ins wörterbuch alle wörter gehören und gleich berechtigt darin sind", steht die Erkenntnis von dem "unermessenen wortvorrat (der deutschen sprache)"<sup>15</sup> gegenüber; jener korrespondiert mit dem Ziel, "dasz das wörterbuch sich über die gesamte hochdeutsche schriftsprache . . . erstrecken solle", diese mit der Feststellung, daß "die menge der in vier jahrhunderten geschriebenen und gedruckten bücher . . . unermeslich (ist)".

1. Für die Berücksichtigung der Quellen folgt daraus, "dasz keinem (der bücher) der zutritt abgeschnitten werde", wenngleich "die unmöglichkeit alle oder nur die meisten (bücher) . . . wirklich vorzuführen . . . am tage (liegt)". Diese Unmöglichkeit ist zusätzlich begründet in der mehr praktischen Erfahrung, daß "alle diese werke (nirgend) vollständig verzeichnet, nicht einmal den geübtesten kennern bekannt, noch weniger irgendwo zusammen aufbewahrt (sind)"<sup>16</sup>: Es ist prinzipiell und praktisch nicht möglich, das Gesamt der vom Objektbereich her infrage kommenden Quellenwerke auch nur annähernd für die Bearbeitung des Wörterbuchs heranzuziehen: Nicht zu erfassende Komplexität des Quellenmaterials.

2. Entsprechend ist es notwendig, eine "ganz beschränkte samlung" von Quellen dem Wörterbuch zugrunde zu legen, die als "verhältnismässig . . . kleiner theil der ausgedehnten deutschen literatur" zu gelten hat. Grimm trifft die Auswahl nach einem praktischen Kriterium, indem er sich weitgehend (vgl. aber auch 1.2.) auf die vorgegebene, zugängliche "eigne . . . samlung" beschränkt: Es ist notwendig, für das Wörterbuch eine beschränkte Quellensammlung vorzusehen: Notwendige Beschränkung des Quellenmaterials.

3. Diese Sammlung "hat gleichwohl den unvermeidlichen einfluss üben müssen, dasz die von ihr selbst dargebotnen, längst gebrauchten und vertrauten ausgaben den besseren vorgezogen wurden, die anderswo zu erlangen gewesen wären."<sup>17</sup> Der hier angesprochene Einfluß der ausgewählten Quellen läßt sich generalisieren: Jede Auswahl von Quellen als Teildokumentation des Objektbereichs begrenzt diesen zunächst rein quantitativ; als Auswahl bestimmter Quellen der unterschiedlichsten Art begrenzt sie ihn aber auch qualitativ. Diese Begrenzung kann bei einer Inkongruenz zwischen dem Auswahlmaterial und dem gesamten Objektbereich zu einer Inkongruenz zwischen Wörterbuch als Beschreibung des Objektbereichs und dem Objektbereich, zu einer Verzerrung und Verfälschung der Ergebnisse durch die Macht des Faktischen, des Vorgegebenen führen.

Entsprechend "kam es" bei der dem Objektbereich adäquaten Auswahl von Quellen "darauf an in jedem Jahrhundert die mächtigsten und gewaltigsten zeugen der sprache zu erfassen und wenigstens ihre grössten werke in das wörterbuch einzutragen", so etwa "KEISERSBERG, LUTHER, HANS SACHS, FISCHART, GÖTHE", so etwa ETTNER, LIEBIG und KANT als Repräsentanten der Sprache der Medizin, Chemie und Philosophie.<sup>18</sup> Kriterien für die Auswahl sind die Verteilung über den gesamten zu dokumentierenden Zeitraum, die besondere Bedeutung einzelner Autoren sowie die Erfassung der Sprachen der verschiedenen Stände und Fachbereiche: Es ist bei der Auswahl der Quellen notwendig, für den Objektbereich repräsentative Werke auszuwählen – repräsentativ bezüglich der zeitlichen Verteilung, der Bedeutung einzelner Autoren und auch verschiedener Fachbereiche: Möglichst weitgehende Repräsentativität des ausgewählten Quellenmaterials.

#### 1.4. Quellenbelege

Die Quellen sollen in Form von Belegen für den einzelnen Wortartikel ausgewertet und verfügbar gemacht werden. Damit ist die Frage nach dem Grad der Vollständigkeit der Exzerption gestellt.

1. Der prinzipiellen und praktischen Unmöglichkeit, alle Quellen heranzuziehen, entspricht die praktische Unmöglichkeit, alle Belege aller Wörter aus den ausgewählten Quellen vollständig zu erfassen: "Wie hätte die ganze anzahl der verzeichneten werke vollständig können gelesen, ausgezogen und eingetragen werden?" Daraus legitimiert sich die Praxis, daß aus "manchen der gebrauchten bücher . . . nur wenige, aus einigen sogar einzelne stellen entnommen worden (sind), wie sie sich zufällig oder auch bei absichtlichem nachschlagen darbieten": Es ist aus praktischen Gründen nicht möglich, das Gesamt aller Belege aller Wörter der ausgewählten Quellen für die Bearbeitung des Wörterbuchs vollständig bereitzustellen: Nicht zu erfassende Komplexität der Quellenbelege.

2. Dem Grundsatz der prinzipiellen Belegbezogenheit des Wörterbuchs steht die Feststellung gegenüber, daß die vollständige Anführung aller Belege zu "einer lästigen und störenden häufung der stellen" führt, was nicht Ziel "der beabsichtigten vollständigkeit" ist und was sich auch durch die vom Werk selbst gesetzten Grenzen verbietet: "der dem wörterbuch gesteckte raum wäre unabsehlich erweitert und ausgedehnt worden". Allerdings werden von Grimm zumindest in der Vorrede konkrete Angaben zum Gesamtumfang des Wörterbuchs und zur Zahl der Einträge nicht gemacht<sup>19</sup>: Es ist notwendig, die Quellenbelege zu den Wörterbucheinträgen der Zahl nach zu beschränken und in eine angemessene Relation zum Gesamtumfang zu bringen: Notwendige Beschränkung der angeführten Quellenbelege.

3. Die Beschränkung der Belege und die damit aufgegebenen Frage ihrer Verteilung auf die einzelnen Stichwörter hat zwei Aspekte zu berücksichtigen: den der Sprache, des jeweiligen Wortes und den der Quelle, des jeweiligen Autors.

Das Ziel der "beabsichtigten vollständigkeit" liegt "in der genauesten ermittelung aller einzelnen wörter . . . die fülle der reichen und herrschenden wörter soll beleuchtet, die unscheinbarkeit der armen und vergessenen unverachtet bleiben." Die zahlenmäßig unterschiedliche Verteilung der Belege im Wörterbuch soll dem unterschiedlichen Grad der Gebrauchshäufigkeit der Wörter entsprechen. "Bei einer menge von wörtern geschah die häufung mit allem bedacht, um keinen zweifel über ihre ausbreitung zu lassen, so wie umgekehrt aus der belege seltenheit die unbeliebtheit eines ausdrucks folgt".

Zweites Kriterium für eine quantitativ unterschiedliche Anführung von Belegen ist die Wichtigkeit bestimmter Quellen, die spezifischen Gebrauchsweisen wichtiger Autoren: "Hin und wieder wird man der Belege zu viel angebracht meinen, namentlich aus LUTHER und GÖTTE. doch jenes einfluss auf die sprache, GÖTTES macht über sie müssen reich und anschaulich vorgeführt werden". So haben zahlreiche Belege etwa bei *ahnungsvoll* oder *bethätigen* den Zweck, "den wachsthum und die befestigung göttischer lieblingswörter recht zu zeigen"<sup>20</sup>: Es ist bei der Auswahl der Belege notwendig, für das einzelne Stichwort repräsentative Belege anzuführen, um sowohl die generelle Gebrauchshäufigkeit als auch die spezifischen Gebrauchsweisen wichtiger Autoren zu dokumentieren: Möglichst weitgehende Repräsentativität der ausgewählten Quellenbelege.

#### 1.5. Nachweis der Belege und Verzeichnis der Quellen

"Wörter verlangen beispiele, die beispiele gewährt, ohne welche ihre bestekraft verloren gieng." Zur "bewährung" der Belege gehört, auch "die quelle (so) anzuführen", daß sie nachschlagbar ist.

1. "unbelegte citate sind unordentlich zusammengefaßte, unbeglaubigte, unbeeidete zeugen." Die bei den einzelnen Wörtern angeführten Beispiele müssen nicht nur mit dem Namen "ihres urhebers", sondern auch mit einer genauen Angabe ihrer "belegstelle" nach Quelle, "nach band und seite" gekennzeichnet sein: denn "sie müssen aufgeschlagen werden können; aus der leichtigkeit dieses nachschlagens entspringt ein großer reiz. . . . der leser hat nicht selten das bedürfnis . . . (die belege) in ihrem vollständigeren zusammenhang einzusehen . . . , wodurch ihm das verständnis vollends erschlossen wird"<sup>21</sup>: Es ist notwendig, die angeführten Quellenbelege mit einer genauen Quellen- und Stellenangabe auszuweisen, um dem Benutzer das Nachschlagen und die Einsicht in den weiteren Kontext zu ermögli-

chen: Notwendiger Nachweis der Quelle und der genauen Stelle angeführter Belege.

2. Entsprechend ist es unentbehrlich, alle "genützten und zugezogenen werke . . . in einem beigefügten verzeichnis" aufzuführen. In der bandweisen Erstellung des Wörterbuches liegt begründet, daß "fortwährend andere schriften hinzutreten", wobei gegenwärtig noch nicht entschieden werden kann, "ob diese in den folgenden bänden jedesmal nachzutragen oder nach beendigung des ganzen wörterbuchs einem umfassenden hauptverzeichnis einzuverleiben sind". Als "dem werke nachtheilig" sieht Grimm an, daß das zunächst gelieferte vorläufige Verzeichnis, "so viel darin enthalten sein mag, die groszen lücken unverdeckt hervortreten" läßt. Resignierend stellt er fest: "es war kein ausweg"<sup>22</sup>: Es ist für die Benutzung des Wörterbuchs unentbehrlich, ein umfassendes Verzeichnis aller herangezogenen Quellen zu erstellen: Notwendiges Verzeichnis aller angeführten Quellen.

## 2. Gesichtspunkte zum Thema 'Korpora und Belegsammlungen für das Interdisziplinäre deutsche Wörterbuch'

Bei den folgenden Überlegungen möchte ich ausgehen von den ersten beiden unter 1.1. und 1.2. erörterten, leicht umformulierten Gesichtspunkten:

- Maschinelles Textkorpus als Datenbasis für das Wörterbuch (vgl. 2.1.)
- Planung der Korpusbereitstellung und Sicherstellung der Hilfeleistung anderer (vgl. 2.2.)

Des weiteren möchte ich die übereinstimmenden Kernbegriffe der Unterpunkte zu 1.3. und 1.4. als zielführende Gliederungspunkte wählen:

- Komplexität (vgl. 2.3.)
- Beschränkung (vgl. 2.4.)
- Repräsentativität (vgl. 2.5.)

Der Punkt "Nachweis der Belege und Verzeichnis der Quellen" wird unter 2.6. aufgenommen.

### 2.1. Maschinelles Textkorpus als Datenbasis für das Wörterbuch

Von dem Charakter des geplanten Unternehmens selbstverständlich (vgl. 0.3.) und von vorneherein vorgesehen war, daß das Wörterbuch nur auf der Grundlage eines maschinell gespeicherten und auswertbaren Textkorpus erarbeitet werden kann und soll.<sup>23</sup>

1. Das Textkorpus, die für die lexikographische Beschreibung des Gebrauchs der Wörter notwendige Dokumentation von Sprache, ist in Art und Zusammensetzung vom Objektbereich des Wörterbuchs her bestimmt.

Dieser ergibt sich aus den Thesen 3 und 20: Die deutsche Standardsprache und ihre Fachsprachen in den deutschsprachigen Ländern und Gebieten mit ihren jeweiligen regionalen Besonderheiten, und zwar von 1945 an bis zur Bearbeitungszeit.<sup>24</sup>

2. Die Notwendigkeit eines maschinell gespeicherten und auswertbaren Korpus<sup>25</sup> ist letztlich begründet in der Größe und in dem Anspruch dieses Wörterbuchs, im einzelnen u.a.

- in der Komplexität des Objektbereichs und der damit gegebenen Komplexität der Textsammlung und der Belege als Primärmaterial
- in der Komplexität der für ein solches Wörterbuch notwendigen sekundären Informationen zu den Einzeltexten und den einzelnen Elementen der Texte
- in der mit dem Einsatz von Computern in die lexikographische Bearbeitung gegebenen Komplexität der Fragestellungen an das Korpus
- in dem Ziel, je nach Benutzerbedürfnissen verschiedene Publikationsvarianten und Präsentationsformen abzurufen
- in dem mit der vorhersagbaren rapiden Ausweitung vor allem der Fachsprachenschatze gegebenen Problem der schleichenden Veralterung des Korpus

## 2.2. Planung der Korpusbereitstellung und Sicherstellung der Hilfeleistungen anderer

Im Zuge der planenden Vorbereitung dieses Wörterbuchs allgemein, der angesetzten sechs Schwerpunkte insgesamt und des Schwerpunktes "Korpora und Belegsammlungen" speziell müssen eine Bestandsaufnahme des Forschungsstandes der einzelnen Bereiche durchgeführt, die jeweiligen Ergebnisse ausgewertet und zu einer Gesamtkonzeption für das Unternehmen verarbeitet werden. Ein Ziel dabei ist, eine Folge von Arbeitsschritten festzulegen, bei der die Zusammenhänge und Abhängigkeiten zwischen den einzelnen Teilschritten berücksichtigt werden.

1. Für die Korpusbereitstellung sind bestimmte Relationen wichtig, u.a.

- die Relation zwischen der Größe des Wortschatzes, der Anzahl der vorgesehenen Stichwörter im Wörterbuch und der Größe des Korpus
- die Relation zwischen der Größe des Wortschatzes der einzelnen Fächer und Bereiche, der Anzahl ihrer Stichwörter im Wörterbuch und dem Anteil der Texte der verschiedenen Fächer und Bereiche im Korpus
- die Relation zwischen der Größe des Korpus und der Länge der einzelnen gespeicherten Texte

Ausgangspunkt entsprechender Überlegungen könnten Schätzungen über Wortschatzumfangs sowie vorhandene Wörterbücher, Korpora und Belegsammlungen der verschiedensten Art sein, wobei die letzteren nicht nur unter dem Gesichtspunkt jeweils gewonnener und für das neue Wörterbuch verwertbarer Erfahrungen betrachtet werden sollten, sondern auch unter dem Aspekt der Möglichkeit, sie in die Materialsammlung zu integrieren und als Material zu nutzen.

2. Ein anderer Aspekt bei der Bereitstellung des Korpus betrifft den Einsatz von Computern. Die Speicherung und Auswertung des Korpus im Dienste der lexikographischen Bearbeitung ist nur auf einer Großrechenanlage möglich; dies macht unter Auswertung vorhandener Erfahrungen u.a. notwendig

- die Entwicklung einer maschinenspezifischen lexikographischen Beschreibungssprache
- die Entwicklung von Verfahren für Zusatzkodierungen zu den Texten und den Textelementen
- die Entwicklung von spezifischen Auswertungsverfahren zur maschinellen Erstellung etwa von Teilmengen mit spezifischen Merkmalsausprägungen, von verschiedenen Publikationsvarianten u.a.m.

3. Schon in der Phase der Vorbereitung ist eine dauerhafte Zusammenarbeit mit Experten und Institutionen anzustreben, die sich mit Fragen der Lexikologie, mit lexikographischen Arbeiten und/oder mit linguistischer Datenverarbeitung befassen. Diesem Bestreben könnte entgegen kommen, daß in verschiedensten Fachbereichen die Notwendigkeit gesehen wird, die jeweilige Terminologie aufzuarbeiten und zum Teil neu zu regeln. Hier könnten gerade von diesem Wörterbuch mit dem Prädikat "interdisziplinär" Impulse ausgehen, die diese verschiedenen Arbeitsvorhaben und -vorgänge aufeinander abstimmen und koordinieren.<sup>26</sup>

### 2.3. Komplexität

Mit *Komplexität* ist ein zentrales Problem des geplanten Wörterbuchs, damit gleichzeitig der zentrale Grund für die Notwendigkeit, ein maschinelles Korpus einzurichten (vgl. 2.1.) und damit wiederum ein zentrales Problem des notwendigen Korpus benannt. Damit ist gesagt, daß die Komplexität auf der Ebene des Wortschatzes bzw. der Wortschatze, der Texte und der Textelemente bzw. der Belege, der Informationen zu den Texten und zu den Textelementen eine wichtige Rolle spielt.

Grimm betont zwar die Unermeßlichkeit des Wortvorrats der deutschen Sprache, doch bezieht er das Problem der Komplexität nicht darauf, daß "ins wörterbuch alle wörter gehören", sondern — entsprechend der Gleich-



setzung von Sprache und Literatur – auf die prinzipielle und praktische Unmöglichkeit, alle Quellen und alle Belege der ausgewählten Quellen zu erfassen und zu berücksichtigen (vgl. 1.3.1. und 1.4.1.).

### 2.3.1. Komplexität der Wortschätze und die Wörterbücher

Die Komplexität der Wortschätze ist sehr komplex. Dies zielt einmal ab auf die Vielzahl der sich immer stärker ausfächernden Fachbereiche mit- samt ihren ausufernden Fachwortschätzen, innerhalb derer immer neue "Sachen", Geräte, Maschinen, Apparate, Verfahren, Sachverhalte und Zu- sammenhänge neu, differenzierter oder umbenannt werden müssen. Zum an- dern ist gemeint, daß sowohl die Gemeinsprache als auch die Fachsprachen in sich nicht homogen, sondern vielfach strukturiert und geschichtet sind.

Wirklich zuverlässige einschlägige Zahlen zu finden ist schwierig, doch ver- mitteln die wenigen, wenn auch zum Teil widersprüchlichen Angaben eine erste Vorstellung von den Dimensionen.

1.1. Der Fachwortschatz der Elektrotechnik (einschließlich der aus meh- reren Wörtern gebildeten Ausdrücke) wird für 1935 auf 25.000 Begriffe angesetzt, eine andere Schätzung spricht für 1930 von 60.000 stehenden Begriffen. Die Berechnung des Wortschatzes der Elektrotechnik auf der Grundlage des 2. Bandes der Illustrierten Technischen Wörterbücher (1928) ergibt 21.000 Ausdrücke. Die übrigen 15 bis 1930 erschienenen Bände der ITW enthalten zusammen 75.000 Wortstellen. Die Bände erfassen folgende Bereiche:

Ersch.- Jahr	Nr.	Titel	Seiten- zahl
1906	1	Die Maschinenelemente und die gebräuchlichsten Werkzeuge	403
1907	2	Die Elektrotechnik	2100
1908	3	Dampfkessel, Dampfmaschinen, Dampfturbinen	1322
	4	Verbrennungsmaschinen	618
1909	5	Eisenbahnbau und Eisenbahnbetrieb	884
	6	Eisenbahnmaschinenwesen	796
1910	7	Hebemaschinen und Transportvorrichtungen	651
	8	Der Eisenbeton im Hoch- und Tiefbau	415
	9	Werkzeugmaschinen	706
	10	Motorfahrzeuge	996
1911	11	Eisenhüttenwesen	785
1915	12	Wasser-, Luft- und Kältetechnik	1959
1919	13	Baukonstruktionen	1030
1923	14	Faserrohstoffe	500
1925	15	Spinnerei und Gespinste	951
	16	Weberei und Gewebe	710
1928	2	Elektrotechnik und Elektrochemie (2. Aufl.)	1304

Der Bestand der Sprache der Chemie wird einmal auf 60.000 bis 100.000 Wörter geschätzt, zum anderen für 1964 auf 2 Millionen. Das elektronische Wörterbuch der Fachsprachen (Russisch/Deutsch/Englisch) der Technischen Universität Dresden umfaßt im Bereich der Chemie 55.000 Begriffsstellen mit 154.000 Lexemen (Englisch-Deutsch), im Bereich der Elektrotechnik 60.000 Begriffsstellen mit 160.000 Lexemen (Russisch-Deutsch). Für die Nachrichten- und Datentechnik mit all ihren Anwendungsgebieten veranschlagte man 1969 rund 1 Million Begriffe. Die Zahl allein der Tier- und Pflanzenarten wird unter Berufung auf den belgischen Biologen De Smeet mit nicht weniger als 60 Millionen angegeben, von denen nur 4 Millionen nicht ausgestorben sind.<sup>27</sup> Aus dem Bereich der Normung des DIN liegen etwa 16.000 Normblätter vor. Die Zahl der in diesen Blättern vorkommenden definierten und undefinierten Termini und Benennungen wird auf 200.000 bis 300.000 geschätzt.<sup>28</sup>

1.2. Die vorstehenden Zahlen illustrieren nur einige Fächer und sind mithin exemplarisch. Die Potenzierung der Komplexität wird daran deutlich, daß eine oft zitierte Schätzung die "Gesamtzahl der Fachgebiete und damit der Fachsprachen" mit rund 300 angibt.<sup>29</sup>

2. Die Zahlen für den Wortschatz der Gemeinsprache muten eher bescheiden an. Der Gesamtwortschatz wird zwischen 300.000 und 500.000 angesetzt, "der Wortschatz der deutschen Standardsprache der Gegenwart einschließlich aller sozialen und regionalen Varianzen . . . und die Teile der Fachwortschätze, die der Standardsprache zugewandt sind und auf sie einwirken", mit etwa 150.000 bis 180.000 Wörtern.<sup>30</sup> Die Größe des aktiven allgemeinsprachlichen Wortschatzes von Gebildeten wird mit rund 10.000 angegeben, die Zahl der von Theodor Storm in seinem gesamten Werk verwendeten verschiedenen Wörter mit 22.500, die Anzahl der im Korpus des Goethe-Wörterbuchs belegten Wörter mit ca. 80.000.<sup>31</sup>

Zu einigen primär gemeinsprachlichen Wörterbüchern liegen folgende Zahlen vor: Die einbändigen Wörterbücher von Wahrig und Mackensen enthalten etwa 100.000 bzw. etwa 145.000 Stichwörter, das sechsbändige Wörterbuch von Klappenbach/Steinitz etwa 85.000; das auf sechs Bände geplante Wörterbuch aus der Duden-Redaktion sieht ca. 160.000 Stichwörter mit ca. 500.000 Definitionen vor.<sup>32</sup>

### 2.3.2. Komplexität der Texte und die Korpora

Der oben (vgl. 2.1.) beschriebene Objektbereich wird dokumentiert in deutschsprachigen schriftlichen und mündlichen Texten der Umgangs- und Alltagssprache, der öffentlichen Verkehrssprache und der Literatursprache sowie der Fachsprachen in den deutschsprachigen Ländern und Gebieten mit ihren regionalen Besonderheiten, und zwar im Zeitraum von 1945 an bis zur Bearbeitungszeit.

1. Der mit den einschlägigen schriftlichen Texten gegebene Umfang wird angedeutet in der folgenden Tabelle der Titelproduktion allein der Bundesrepublik Deutschland einschließlich Berlin (West) von 1951 bis 1976:

Titelproduktion der Bundesrepublik Deutschland  
einschl. Berlin (West) 1951 bis 1976

Titel				
Jahr	Insgesamt	Erstauflage	Neuauflage	Erstauflage: Neuauflage
1951	14 094	10 149	3 945	72:28
1952	13 913	10 536	3 377	76:24
1953	15 738	12 096	3 642	77:23
1954	16 240	12 264	3 976	76:24
1955	16 660	12 701	3 959	76:24
1956	17 215	13 307	3 908	77:23
1957	16 690	13 017	3 673	78:22
1958	20 476	16 099	4 377	79:21
1959	16 532	12 725	3 807	77:23
1960	22 524	17 639	4 885	78:22
1961	23 132	17 717	5 415	77:23
1962	22 615	17 854	4 761	79:21
1963	25 673	20 553	5 120	80:20
1964	26 228	20 940	5 288	80:20
1965	27 247	22 842	4 405	84:16
1966	23 777	20 169	3 608	85:15
1967	30 683	25 235	5 448	82:18
1968	32 352	27 293	5 059	84:16
1969	35 577	29 929	5 648	84:16
1970	47 096	38 703	8 393	82:18
1971	42 957	36 116	6 841	84:16
1972/1973	93 498	79 641	13 857	85:15
1974	49 761	41 276	8 485	83:17
1975	43 649	35 486	8 163	81:19
1976	46 763	37 628	9 135	80:20
1951-1976	741 090	601 915	139 175	81:19

Aus der Tabelle folgt zunächst, "daß sich die Zahlen der jährlichen Veröffentlichungen in den letzten Jahren gegenüber dem . . . Ausgangsjahr (1951) verdreifacht haben. Allein auf die Erstauflagen bezogen, wurden im Durchschnitt der Jahre seit 1970 pro Jahr viermal soviele Titel vorgelegt wie 1951. Die Zahl der Neuauflagen war in dem genannten Zeitraum etwa doppelt so groß."

Selbst wenn man bedenkt, daß unter den produzierten Titeln auch Neuauflagen in der Spanne von 28 bis 15% enthalten sind, deren Erstauflage möglicherweise vor 1945 liegt<sup>33</sup>, und daß die Anteile der Übersetzungen ins Deutsche "im Zeitabschnitt zwischen 1962 und 1976 . . . in der Spanne zwischen 9,2 v.H. und 12,5 v.H. (liegen)"<sup>34</sup> und man zumindest diskutieren kann, ob diese Titel den oben genannten Merkmalen entsprechen<sup>35</sup>; selbst wenn man zusätzlich bedenkt, daß etwa 20 bis 25% der Titel umfangmäßig unter 49 Seiten liegen<sup>33</sup> und daß bestimmte Arten von Titeln wie etwa Wörterbücher oder Lexika anders anzusehen sind als Werke mit zusammenhängenden Texten, so ist die Menge der einschlägigen Titel und ihr Umfang immens groß; dies vor allem auch unter dem Gesichtspunkt, daß hier nur über Buchtitel der Bundesrepublik (einschließlich Berlin-West) berichtet wird, und nicht über Titel deutschsprachiger Bücher in der Schweiz, in Österreich, in der DDR u.a., über schriftliche Texte, die nicht Buchform haben wie Zeitungen, und über den großen Bereich gesprochener Texte.

2. Dokumentationen aus diesem Textbereich liegen in Form verschiedenster Textkorpora vor. Schaefer<sup>36</sup> hat eine Übersicht von zehn bzw. dreizehn maschinenlesbaren Korpora mit mindestens 200.000 Wortstellen vorgelegt, deren deutschsprachige Texte bezüglich der von Schaefer gewählten Merkmale wie etwa national, medial, diastratisch recht unterschiedlich sind. Die Gesamtzahl der Wortstellen der Korpora insgesamt liegt bei ca. 18 Millionen. Auf weitere Korpora mit wissenschaftlichen, medizinischen Texten und solchen aus dem Bereich des Steuerrechts sei hier nur hingewiesen.<sup>37</sup>

### 2.3.3. Komplexität der Belege und die Belegsammlungen

Aus maschinell verfügbaren Korpora etwa der oben genannten Art können systematisch alle Belege aller vorkommenden Wörter mit Umtext gewonnen werden – im Unterschied etwa zu den Zeiten Grimms (vgl. 1.4.1.).

1. Mit dieser technischen Möglichkeit stellt sich das Problem der Komplexität insofern, als auf der einen Seite Lexeme übermäßig häufig belegt sind, während andere nur selten, nur einmal oder überhaupt nicht vorkommen:

- So kommen etwa in einem Korpus von 3 Millionen Textwörtern 120.000 verschiedene Wortformen vor, von denen 40.000 (33%) nur einmal belegt sind. In einem anderen Korpus mit 7.222.000 Textwörtern und 132.800 Wortformen sind 59.900 Wortformen (45%) einmal, 18.800 (18%) zweimal belegt. Von den 90.400 Lexemen sind 43.000 (48%) einmal, 12.400 (14%) zweimal belegt.<sup>38</sup>
- So ergibt sich bei einem Korpus von 6.890 Belegen mit 1.228 verschiedenen Wortformen folgende Verteilung<sup>39</sup>:

Belegzahl	Wortformen	% von 1.228	% von 6.890
1	702	57,17	10,04
2	179	14,57	5,13
3	86	7,00	3,69
4 – 10	169	13,76	14,52
11 – 326	92	7,5	66,62

- So ergibt sich bei einer Gesamtzahl von 12.464 Belegen mit 1.685 Voll-  
verben folgende Verteilung<sup>40</sup>:

Belegzahl	Verben	% von 1.685	% von 12.464
1	777	46,11	6,23
2	273	16,02	4,38
3	156	9,25	3,75
4 – 10	341	20,23	16,08
11 – 2.338	138	8,39	69,56

- So ergibt sich für das Korpus des Goethe-Wörterbuchs, daß 60% der  
Wörter des Buchstabens A ein- bis dreimal belegt sind.<sup>41</sup>

Hinzu kommt, daß die meisten der obersten Ränge in den Häufigkeitslisten  
von Funktionswörtern, Pronomen oder Artikelwörtern eingenommen wer-  
den, und zwar in den verschiedensten Wortschätzen, wie die folgende Tabel-  
le über englische Fachwortschätze zeigt<sup>42</sup>:

Rang	Bauwesen	Chemie	Mathematik	Medizin	Physik
1.	<i>of</i> 5,3	<i>of</i> 5,5	<i>of</i> 6,5	<i>of</i> 6,4	<i>of</i> 6,4
2.	<i>and</i> 3,1	<i>be</i> 4,6	<i>be</i> 5,0	<i>in</i> 3,7	<i>and</i> 3,5
3.	<i>in</i> 2,8	<i>in</i> 2,9	<i>and</i> 3,6	<i>and</i> 3,2	<i>in</i> 3,2
4.	<i>be</i> 2,5	<i>and</i> 2,8	<i>in</i> 2,8	<i>be</i> 2,9	<i>be</i> 2,6
Sa	13,7	15,8	17,9	16,2	15,7
5.	<i>for</i> 1,3	<i>to</i> 2,0	<i>that</i> 1,9	<i>to</i> 1,6	<i>to</i> 1,5
6.	<i>to</i> 1,3	<i>by</i> 0,9	<i>we</i> 1,9	<i>with</i> 1,0	<i>that</i> 1,3
7.	<i>that</i> 0,9	<i>have</i> 0,8	<i>for</i> 1,6	<i>or</i> 1,0	<i>by</i> 1,3
8.	<i>as</i> 0,8	<i>with</i> 0,8	<i>by</i> 1,3	<i>by</i> 1,0	<i>for</i> 1,2
9.	<i>it</i> 0,8	<i>as</i> 0,8	<i>to</i> 1,2	<i>this</i> 0,9	<i>this</i> 1,0
10.	<i>by</i> 0,7	<i>for</i> 0,8	<i>this</i> 0,9	<i>that</i> 0,8	<i>with</i> 1,0
Sa	19,5	21,9	26,7	22,5	23,0

In diesen auf der Grundlage einer Stichprobe im Umfang von 35.000 erstellten Ranglisten werden die ersten vier Positionen von *of*, *and*, *in* und *be* eingenommen; die Textdeckung als Summe der angegebenen relativen Häufigkeit bewegt sich zwischen 13,7% und 17,9%.

Betrachtet man die ersten zehn Positionen, so kommen neben den oben genannten Elementen noch *to* und *by* in allen Wortschätzen vor. D.h., von den zehn häufigsten Elementen mit einer relativen Häufigkeit und damit Textdeckung zwischen 19,5% und 26,7% sind sechs in allen fünf Wortschätzen vertreten. Zwei sind in vier (*for* und *that*) bzw. in drei (*with* und *this*) vertreten, *as* in zwei und *it*, *we*, *have* und *or* in je einem. Außer *be* und *have* gehören alle Elemente dieser zehn Ränge zu der oben charakterisierten Gruppe von Wörtern, deren Belegzahlen sicher nicht in direkter, sondern eher in reziproker Relation zur Wörterbuchrelevanz stehen.

2. Neben den Textkorpora als Speicher auch für abrufbare Belege können manuell oder photomechanisch erstellte Belegsammlungen mitherrangezogen werden, allerdings wohl mehr als komplementäres Korrektiv in flankierender Funktion, da sie zumeist durch gezielte Exzerption entstanden und weder maschinell noch statistisch auswertbar sind.

An vorhandenen Belegsammlungen sind etwa zu nennen:

- Fremdwortkartei des Instituts für deutsche Sprache: 2 Millionen Belege auf Zetteln (manuell); 14. Jahrhundert bis 1972; Anwendung: Wortgeschichte und Belege für die Zeit nach 1945<sup>43</sup>
- Dudenkartei: ca. 5 Millionen Belege auf DIN-A-6-Karten (manuell); Stamm- und Verweiskartei; zum geringeren Teil vor 1945; Anwendung: Belege für die Zeit nach 1945<sup>44</sup>
- Kartei des Deutschen Wörterbuchs (Grimm)
  - <sup>1</sup> DWB (letzte Lieferung 1960): ca. 2 Millionen Belege im Wörterbuch ausgedruckt
  - <sup>2</sup> DWB (Göttingen): D – F 2,7 Millionen Belege auf DIN-A-6-Zetteln (photomechanisch); Anwendung: Wortgeschichte<sup>45</sup>
- Korpus des Goethe-Wörterbuchs: 3,2 Millionen Belege auf DIN-A-6-Zetteln (manuell, zum geringeren Teil maschinell); Anwendung: Wortgeschichte<sup>46</sup>
- Sammlung des Rechtswörterbuchs (Heidelberg): 2,5 Millionen Belege auf Zetteln; bis 1750 (zum Teil 1832); Anwendung: Wortgeschichte<sup>47</sup>

### 2.3.4. Komplexität: Reduktion und Repräsentativität

Die in den vorstehenden Abschnitten angedeutete Komplexität der Wortschätze, Texte und (Text-)Belege spiegelt sich wider in der großen Zahl vorhandener Wörterbücher, Textkorpora und Belegsammlungen. Diese sind als Versuche, sprachliche Realität abzubilden, zu dokumentieren, zu beschreiben und reflektiert zu vermitteln, eine modellhafte Reduktion unter einem jeweils spezifischen Aspekt oder einem spezifischen Bündel von Aspekten – Reduktion als verkürzende Beschränkung von etwas auf etwas.

Ein Aspekt der Komplexität ist die quantitative Vielzahl, die Menge, der Umfang; ein anderer Aspekt ist die qualitative Vielfalt, die Schichtung, die Auffächerung, die – vor allem, wenn als Problem empfunden – gern wohl auch *Heterogenität* genannt wird.

Die qualitative Vielfalt ist zunächst ein Problem der Beschreibung, denn das qualitativ Unterschiedliche erfordert eine je spezifische Darstellungsweise. Unter dem Zeichen der Notwendigkeit einer quantitativen Beschränkung – gezeichnet in diesem Sinne sind alle Wörterbuch-, Korpus- und Belegsammlungenunternehmen –, unter dem Zwang der mengenmäßigen Reduktion wird die qualitative Vielfalt zudem zu einem Problem der quantitativen Verteilung, das nicht bestünde, wenn alles erfaßt werden könnte. Überlegungen zur Lösung dieses Verteilungsproblems bilden den Kontext zur Verwendung des Ausdrucks *Repräsentativität*.<sup>48</sup>

### 2.4. Beschränkung

Angesichts der Weite des Objektbereichs in all seiner Komplexität ist für das Interdisziplinäre deutsche Wörterbuch eine Reduktion unabdingbar, auch dann, wenn es auf 20 Bände geplant ist (These 2).

#### 2.4.1. Anzahl der Einträge – Größe des Korpus

Ausgehend von Stichwortzahlen vorhandener Wörterbücher und Lexika und der Größe vorhandener Korpora, ausgehend von der Berechnung, daß ein Korpus als Grundlage des Wahrigschen Wörterbuchs Texte etwa mit 10 bis 12 Millionen Textwörtern enthalten haben müßte, wurde auf Vorschlag von Engelen die Zahl der Wörterbuchartikel, der Lemmata auf mindestens 200.000 und der Umfang des Korpus auf mindestens 50 Millionen Textwörter festgelegt (These 18).<sup>49</sup> Beide Angaben sind als Minimalgrenzen anzusehen.

Als Maximum für die Länge des einzelnen Textes wurde die Zahl von 10.000 Textwörtern angegeben, wodurch die Zahl der Texte auf mindestens 5.000 festgelegt wird. Als oberste Grenze der bei der Bearbeitung eines Stichworts zu berücksichtigenden Belegzahl wurde 2.000 genannt.

#### 2.4.2. Rückgriff auf vorhandene Materialien

Aus praktischen und ökonomischen Gründen liegt der Gedanke nahe, auf die vorhandenen Wörterbücher, Korpora und Belegsammlungen zurückzugreifen, so wie Grimm die "eigne ganz beschränkte samlung" von Quellen seinem Wörterbuch zugrunde legte.

1. Vorhandene Wörterbücher müßten einmal auf ihre lexikographische Praxis hin ausgewertet werden, doch stimmt es nicht besonders hoffungsvoll, wenn Weinrich feststellt, "daß der deutsche Sprachraum, insbesondere die Bundesrepublik, im 20. Jahrhundert kein gutes Wörterbuchland mehr ist" und daß die vorhandenen "Fachwörterbücher, gleich welcher Qualität, . . . unter linguistisch-lexikographischen Gesichtspunkten Wildwuchs (sind)".<sup>50</sup>

Zum anderen könnten die Einträge vorhandener ein- und mehrsprachiger Wörterbücher und Lexika besonders auch der Fachsprachen maschinell aufgelistet, nach ihrer Häufigkeit in den verschiedensten Wörterbüchern ausgewertet und zu einer ersten vorläufigen Stichwortliste allgemein und für die einzelnen Fächer zusammengestellt werden. Diese Liste könnte mit der aus dem Gesamtkorpus erstellten Wortliste verglichen bzw. vereinigt werden.<sup>51</sup> Bedenken gegen solch ein Verfahren ergeben sich daraus, daß Erhebungen dieser Art sich nur auf die Ausdrucksseite beziehen und daß linguistische Folgerungen aus diesen statistischen Erhebungen nur schwer zu ziehen sind (Engelien).

2. Vorhandene Korpora wären zunächst etwa nach folgenden Gesichtspunkten zu prüfen:

- Systematik der Speicherung
- Verfügbarkeit und Kompatibilität
- Texte nach 1945
- Grenze für den Einzeltext: 10.000 Textwörter
- Übersetzungen und Neuauflagen

Wenn man auf die von Schaefer erfaßten 13 Korpora mit ca. 18 Millionen Textwörtern diese Gesichtspunkte anwendet, so ergibt sich, daß die Texte aller Korpora mit wenigen Ausnahmen in die Zeit nach 1945 fallen, daß aber bezüglich der Systematik der Erfassung und der Länge der Einzeltexte gravierende Unterschiede bestehen: Während etwa das Mannheimer Korpus nur vollständige Texte enthält, erfaßt etwa das LIMAS-Korpus mit 2.000 Wörtern pro Text auch Teile aus längeren Texten. Während bei bestimmten Korpora die Textlänge aller Texte gleich ist, gibt es bei anderen Korpora sehr starke Schwankungen. Während die Textlänge bei fünf Korpora



nicht größer als 10.000 Textwörter ist, überschreitet sie bei vier Korpora diese Obergrenze erheblich; für die restlichen vier Korpora fehlen entsprechende Angaben. Spezifisch für das Bonner Korpus ist eine gleichmäßige Stufung der zeitlichen Herkunft seiner Texte, spezifisch etwa für das Aachener, das Bonner und das Saarbrücker Korpus die Berücksichtigung auch von Übersetzungen.

Nimmt man zu diesen schon erheblichen Unterschieden hinzu

- daß für die vorhandenen Korpora kaum Auskünfte über den Grund des gewählten Umfangs vorliegen
- daß den Korpora sehr differierende Zweckbestimmungen und unterschiedliche Klassifikationssysteme zugrundeliegen, deren Kategorien nicht selten "erst nach Fertigstellung des Korpus ausgewählt worden (sind), um eine Textauswahl nachträglich beschreiben, rechtfertigen oder auf ihre Relevanz hin überprüfen zu können"<sup>52</sup>
- daß die Berücksichtigung von Texten mit Ausprägungen etwa der Merkmale national, diastratisch und Genre sehr variabel ist

dann ist sehr zweifelhaft, ob eine erste Beschränkung auf diese Korpora als Grundstock für das Gesamtkorpus praktisch und sinnvoll ist.

3. Gegenüber den vorhandenen Belegsammlungen sind ähnliche Bedenken, wenn auch in noch stärkerem Maße wie gegenüber den Korpora anzumelden, zumal vier der oben genannten fünf Sammlungen allenfalls für die historische Komponente des Wörterbuchs<sup>24</sup> verwendbar sind.

#### 2.4.3. Verzerrung durch die Macht des faktisch Vorgegebenen

Es zeigt sich, daß der unter 2.4.1. zahlenmäßig abgesteckte und damit beschränkte Rahmen des Interdisziplinären deutschen Wörterbuchs und seines Korpus weder vorläufig noch zunächst zum Teil unter Rückgriff auf vorhandene Materialien ausgefüllt werden kann. Diese begrenzte, in ihrem Konzept nicht aufeinander abgestimmte und damit zu heterogene Ansammlung von Auswahlmaterialien birgt die Gefahr in sich, daß das Wörterbuch durch die Macht des faktisch Vorgegebenen zu einer verfälschten und verzerrten Beschreibung des Objektbereiches wird. Das schließt natürlich nicht aus, daß einzelne Korpora oder Teile verschiedener Korpora bei der endgültigen Zusammenstellung des Gesamtkorpus berücksichtigt werden können und sollten.

#### 2.5. Repräsentativität

Im folgenden sollen einige Überlegungen angestellt werden zur Repräsentativität des Korpus als Auswahl aus der Gesamtheit der Texte, die mit Kennzeichnung des Objektbereichs des Interdisziplinären deutschen Wör-

terbuchs beschrieben sind. "Es wäre die definierte Menge aller individuellen Sprachrealisationen dann angemessen repräsentiert, wenn entsprechend der Häufigkeit ihres festgestellten Vorkommens bzw. Gebrauchs die jeweiligen Teilmengen aufgrund bestimmter statistischer Kriterien durch entsprechend ausgewählte Textexemplare angemessen vertreten wären."<sup>53</sup>

Für die Zusammenstellung des Korpus sind bestimmte, oft dichotomische Gesichtspunkte zu berücksichtigen, von denen einige im folgenden angesprochen werden sollen.

### 2.5.1. Temporal und national/areal

1. Das Korpus soll sich auf Texte von 1945 bis zur Bearbeitungszeit beschränken (These 20). Wenn man unter Ansatz einer bestimmten Vorbereitungszeit als Beginn der eigentlichen Bearbeitung ca. 1985 ansetzt, dann stellt sich die Frage nach der zeitlichen Verteilung der Texte. Gegenüber einer wahllos-unsystematischen Streuung über diesen Zeitraum hin besteht die Möglichkeit, die oben angesetzte Mindestzahl von 5.000 Texten mit je 10.000 Textwörtern gestaffelt zu verteilen:

Schnitt	Stufen	Zahl der Texte pro Stufe	Zahl der Textwörter pro Stufe
1945, 46 ... 84	40	125	1.250.000
1945, 50 ... 80	8	ca. 625	6.250.000
1945, 55 ... 85	5	1.000	10.000.000
1945 – 1949 1955 – 1959 1965 – 1969 1975 – 1979	4	1.250	12.500.000

2. Neben dem temporalen ist der areale Aspekt bei der Textdokumentation insofern zu berücksichtigen, als nicht nur Texte aus der Bundesrepublik, sondern zumindest auch aus der DDR, Österreich und der Schweiz erfaßt sein müssen. Der Schlüssel für die Verteilung könnte die Titelproduktion der einzelnen Länder sein (für 1968: BRD 30.323, DDR 5.568, Österreich 5.164, Schweiz 6.228), wobei der Anteil der DDR relativ klein ist und zumindest in den Jahren von 1963 bis 1970 relativ konstant in der Spanne von 5.748 und 5.169 geblieben ist.<sup>54</sup>

### 2.5.2. Geschrieben – gesprochen

Das Korpus soll nicht nur Texte der geschriebenen, sondern auch der gesprochenen Sprache enthalten, und zwar nicht nur der Standardsprache, sondern auch der Fachsprachen (These 15). Angesichts des hohen Zeit- und Personalaufwandes für die Aufbereitung entsprechender Texte sollte

der Anteil am Gesamtkorpus möglichst gering gehalten werden. Grundstock für diese Komponente könnte das Freiburger Korpus der gesprochenen Sprache sein, das aus ca. 220 Texten aus der Zeit von 1967 bis 1972 mit ca. 600.000 Textwörtern besteht. Fünf dieser Texte liegen umfangmäßig über der Grenze von 10.000 Textwörtern, sonst liegt der Umfang zwischen 154 und 9.725 Wörtern.<sup>55</sup>

Überlegungen zur Erweiterung dieses Korpus müßten vor allem den zeitlichen und den fachsprachlichen, daneben aber wohl auch den arealen Gesichtspunkt berücksichtigen. Grundsätzliche Schwierigkeiten des Einbezugs dieses Korpus in ein Gesamtkorpus vor allem geschriebener Sprache ergeben sich daraus, daß die "Freiburger Konzeption von (außerlinguistischen) Redekonstellationstypen, denen je eine Textsorte entspricht, . . . aus einem Modell zur Beschreibung kommunikativer Situationen" gewonnen ist, das auf Texte der geschriebenen Sprache kaum anwendbar sein dürfte.<sup>56</sup>

### 2.5.3. Gemeinsprache – Fachsprachen

Der spezifische Charakter des Interdisziplinären deutschen Wörterbuchs liegt in der starken fachsprachlichen Komponente: Gegenstand ist die "deutsche Standardsprache und ihre Fachsprachen" (These 3 und 12). Schon die wenigen Zahlen oben unter 2.3.1. haben deutlich gemacht, daß eine Reduktion der Gesamtmenge unerläßlich ist. Diese Reduktion setzt abgrenzbare Teilmengen der Gesamtmenge voraus, von denen dann nach bestimmten Kriterien bestimmte ausgegrenzt werden können.

#### 2.5.3.1. Zahl und Klassifizierung der Fachbereiche

Ein erster Schritt könnte darin bestehen, die Fachbereiche zusammenzustellen und – zumindest vorläufig (vgl. These 17) – zu klassifizieren. Doch ist der Entschluß dazu leichter als die Durchführung, denn das "System der Wissenschaften ist komplex und dynamisch; es entwickelt sich unter verschiedenen Gesichtspunkten. Dadurch wird das Auffinden geeigneter einfacher klassifikatorischer Eigenschaften zu einem Problem, das bisher noch nicht gelöst wurde".<sup>57</sup>

Diese Erkenntnis wird sehr schnell bestätigt, wenn man sich verschiedene Versuche der Klassifizierung vor Augen führt:

Kaeding setzt elf verschiedene Bereiche an, auf die sich die 11 Millionen Textwörter seines Korpus wie folgt verteilen<sup>58</sup>:

Gebiet	Prozent	Textwörter
Recht	3,7	407.000
Religion	3,7	407.000
Handel	2,8	308.000
Medizin	0,9	99.000
Geschichte	4,6	506.000
Schöne Literatur	18,3	2.013.000
Korrespondenz	9,2	1.012.000
Politik	8,2	902.000
Naturwissenschaft	3,7	407.000
Kriegswesen	5,5	605.000
Verschiedenes	39,4	4.334.000
Zusammen:	100,0	11.000.000

Zur Klassifizierung der in einer größeren Buchhandlung vorhandenen Fachwörterbücher wurden 24 Fächer unterschieden<sup>59</sup>:

FACHWÖRTERBÜCHER IM BUCHHANDEL		
Fachsprache	Zahl der Fachwörterbücher – Bände	Seiten
1. Theologie	72	60.630
2. Wirtschaft	68	45.912
3. Literatur	50	34.335
4. Geschichte	47	29.035
5. Technik	51	28.510
6. Medizin	45	26.707
7. Soziologie	30	23.020
8. Pädagogik	47	18.744
9. Psychologie	32	18.470
10. Biologie	41	17.450
11. Geowissenschaften/Ethnologie	33	16.110
12. Philosophie	22	15.475
13. Politik	31	14.870
14. Recht	38	14.415
15. Physik	29	12.680
16. Linguistik	37	12.080
17. Handwerk	22	12.033
18. Kunst/Architektur	20	8.078
19. Musik, Theater, Film	23	7.180
20. Chemie	10	5.720
21. Mathematik	14	5.450
22. Sport	15	4.890
23. Hauswirtschaft	8	3.262
24. EDV/Kybernetik	10	3.210

Zur Klassifizierung der Titelproduktion der Bundesrepublik Deutschland sind 25 verschiedene Sparten eingerichtet worden<sup>60</sup>:

Titelproduktion der Bundesrepublik Deutschland einschl. Berlin (West)  
1976 nach Sachgebieten

Sachgebiet	Erst- auflage	Neu- auflage	Ins- gesamt	v.H.
Allgemeines, Buch und Schrift, Hochschulen	559	83	642	1,4
Religion, Theologie	1 637	499	2 136	4,6
Philosophie, Psychologie	959	316	1 275	2,7
Recht, Verwaltung	1 873	721	2 594	5,5
Wirtschafts- und Sozialwissenschaften, Statistik	3 161	588	3 749	8,0
Politik, Wehrwesen	809	130	939	2,0
Sprach- und Literaturwissenschaft	1 391	424	1 815	3,9
Schöne Literatur	8 579	1 271	9 850	21,1
Jugendschriften	2 159	656	2 815	6,0
Erziehung, Unterricht, Jugendpflege	1 837	460	2 297	4,9
Schulbücher	1 498	876	2 374	5,1
Bildende Kunst, Kunstgewerbe	1 371	300	1 671	3,6
Musik, Tanz, Theater, Film, Rundfunk	525	100	625	1,3
Geschichte, Kulturgeschichte, Volkskunde	1 537	322	1 859	4,0
Erd- und Völkerkunde, Reisen	679	221	900	1,9
Karten, Kartenwerke	2 172	673	2 845	6,1
Medizin	1 350	417	1 767	3,8
Naturwissenschaften	1 541	275	1 816	3,9
Mathematik	647	84	731	1,6
Technik, Industrie, Gewerbe	1 791	318	2 109	4,5
Verkehr	342	77	419	0,9
Land- und Forstwirtschaft, Hauswirtschaft	738	221	959	2,0
Turnen, Sport, Spiele	326	100	426	0,9
Verschiedenes	56	3	59	0,1
Kalender und Almanache	91	--	91	0,2
Insgesamt	37 628	9 135	46 763	100,0

Erk gliedert die von ihm erfaßten 34 Fächer in folgende Gruppen<sup>61</sup>:

Grundwissen- schaften	Naturwissen- schaftliche Fächer	Sozialwissen- schaftliche Fächer	Geisteswissen- schaftliche Fächer
Kybernetik	Anthropologie (biol.)	Betriebswirtschaft	Archäologie
Mathematik	Astronomie	Ethnologie (Anthropologie)	Kulturosoziologie
Naturphiloso- phie-Logik	Biologie	Geschichte	Kunstwissenschaft
Statistik	Chemie	Pädagogik	Literaturwissen- schaft
	Geographie	Politische Wissensch.	Musikwissenschaft
	Geologie	Publizistik	Philosophie
	Medizin	Rechtswissenschaft	Religionswissen- schaft
	Meteorologie	Sozialpsychologie	Sprachwissenschaft
	Physik	Soziologie	Theaterwissenschaft
	Psychologie	Wirtschafts- wissenschaft	Theologie

In einem ersten Versuch, die fachsprachliche Komponente des Interdiszi-  
plinären deutschen Wörterbuchs zu gliedern, wurden 62 Bereiche in folgen-  
des Schema gebracht<sup>59</sup>:

I.	Jurisprudenz, Politik, Verwaltung
II.	Wirtschaft und Wirtschaftswissenschaften (Handel, Waren, Produkte, Wer- bung, Verkehr, Banken, Finanzen, Ernährung, Landwirtschaft, Forstwirt- schaft, Gartenbau . . .)
III.	Biologie, Anthropologie, Medizin, Veterinärmedizin, Genetik, Physiologie, Pharmazie
IV.	Philosophie, Mathematik, Logik, Linguistik, Informatik, Kybernetik
V.	Chemie, Physik, Astronomie, Geophysik
VI.	Technik und Industrie (Bauingenieurwesen, Bergbau und Hüttenwesen, Maschinenbau, Elektrotechnik, Feinwerktechnik, Metallbearbeitung und -verarbeitung, Werkstoffe, Handwerk)

VII.	Kultur (Literatur, Kunst, Architektur, Archäologie, Musik, Theater, Film, Fernsehen, Journalismus, Verlagswesen)
VIII.	Psychologie, Pädagogik, Soziologie
IX.	Geschichte mit Theologie (eventuell einschließlich Onomastikon)
X.	Kulturgeographie mit Geologie, Paläontologie, Mineralogie, Ethnologie, Meteorologie, Städtebau, Landesplanung, Raumordnung (eventuell einschließlich Onomastikon)

In der in Bibliotheken, Bibliographien u.ä. häufig verwendeten internationalen Dezimalklassifikation werden zunächst zehn Hauptabteilungen unterschieden und mit den Zahlen 0 bis 9 bezeichnet; jede dieser Hauptabteilungen wird durch Hinzufügung einer zweiten Ziffer in weitere Abteilungen zweiter Ordnung zerlegt usw. Im folgenden werden die zehn Hauptabteilungen und die Untergliederungen auf der zweiten Ebene der dritten, fünften und sechsten Hauptabteilung zusammengestellt <sup>62</sup> :

#### ÜBERSICHT DER HAUPTABTEILUNGEN

0	Allgemeines
1	Philosophie
2	Religion, Theologie
3	Sozialwissenschaften
30	Soziologie, Soziographie
31	Statistik
32	Politik, Allgemeine Staatslehre
33	Volkswirtschaft
34	Recht, Rechtswissenschaft
35	Öffentliche Verwaltung, Kriegskunst, Kriegswissenschaft.
36	Fürsorge, Volkshilfe, Versicherung
37	Erziehung, Unterricht
38	Handel, Verkehr
39	Völkerkunde, Volkskunde, Sitte, Brauchtum, Volksleben
4	Sprachwissenschaft, Philologie
5	Mathematik, Naturwissenschaften
51	Mathematik
52	Astronomie, Geodäsie
53	Physik
54	Chemie, Kristallographie, Mineralogie
55	Geologie, Meteorologie
56	Paläontologie
57	Biologische Wissenschaften, Anthropologie
58	Botanik
59	Zoologie

6	Angewandte Wissenschaften, Medizin, Technik
60	Gemeinsame Fragen der angewandten Wissenschaften, Erfindungen
61	Medizin
62	Ingenieurwesen, Technik
63	Landwirtschaft, Forstwirtschaft, Tierzucht, Jagd, Fischerei
64	Hauswirtschaft
65	Handels- und Verkehrstechnik (Organisation und Betrieb)
66	Chemische und verwandte Industrien
67/68	Verschiedene Industrien und Gewerbe
69	Bauarbeiten, Bauhandwerk, Baustoffe, Hausbau, Wohnungsbau
7	Raumordnung, Landesplanung, Städtebau, Landschaftsgestaltung, Gartenarchitektur
8	Literaturwissenschaft
9	Geographie, Biographien, Geschichte

In einer Aufstellung aus der Dudenredaktion "Bereiche – Fach- und Sondersprachen"<sup>63</sup> werden ca. 170 Bereiche unterschieden, die sich leicht ergänzen lassen zu einer erweiterten Liste von ca. 230 Bereichen. Wüster schätzt die Zahl der Fachgebiete und damit der Fachsprachen auf rund 300.<sup>29</sup> Der Fächerkatalog des Hochschulverbandes enthält 88 Fachrichtungen, 779 Fächer und 2.268 Fachgebiete (Inhalt der Fächer).<sup>64</sup>

Sowohl die sehr unterschiedlichen Zahlen der Fächer als auch die unterschiedlichen Arten der Subklassifizierung zeigen deutlich die Subjektivität solcher Ansätze.<sup>65</sup>

Wenn man den Vorschlag, im Korpus für gemeinsprachliche und fachsprachliche Texte ein Verhältnis von 2 : 3 vorzusehen<sup>66</sup>, aufgreift, so ergäbe sich bei einer Gesamtmenge von mindestens 3.000 Fachtexten à 10.000 Textwörter bei den verschiedenen Ansätzen der Fächer folgende Verteilung:

Fächerzahl	Texte pro Fach	Textwörter pro Fach
10	300	3.000.000
25	120	1.200.000
50	60	600.000
100	30	300.000
300	10	100.000
1.000	3	30.000

Diese Aufstellung ist natürlich sehr mechanisch und berücksichtigt nicht, daß die Fachbereiche verschieden große und unterschiedlich strukturierte Sprachen haben. Davon wird weiter unten noch zu sprechen sein.



### 2.5.3.2. Auswahl bestimmter Fachbereiche und Fachsprachen

Für die Auswahl bestimmter, wichtiger Fachsprachen könnte auf der Grundlage der Klassifikation folgender erster Katalog von Kriterien diskutiert werden: Bezugspunkt für die Bewertung der Wichtigkeit sollte der Nichtfachmann, der Laie mit seinem mehr allgemeinen freiwilligen Interesse an oder in seinem mehr erzwungenen Konfrontiertsein mit verschiedenen Fachbereichen und ihren Sprachen sein (vgl. These 11 und 15):

- Die wirtschaftliche Bedeutung eines Fachbereiches läßt sich zunächst messen am Umsatz, an der Produktstückzahl sowie an den Lohn- und Gehaltskosten. Der Umsatz ist möglicherweise Index für die mehr allgemein wirtschaftliche Bedeutung. Die Lohn- und Gehaltskosten korrespondieren mit der Zahl der im Verbundsystem eines Faches Beschäftigten und dadurch mit dem Grad der allgemeinen Verflechtung eines Faches mit dem Arbeitsbereich der Menschen. Die Produktzahl ist die Grundlage für den Produktverkauf und die Produktstreuung. Diese korrespondiert mit dem Grad der allgemeinen Verflechtung eines Faches mit dem Privatbereich der Menschen. Zu prüfen wäre, ob die Unterscheidung in 'Grundprodukte, Konsumprodukte und Luxusprodukte' mit der Kaufmotivation 'gezwungenermaßen (weil notwendig fürs Leben), teils gezwungen – teils freiwillig (weil üblich) und relativ freiwillig (weil überflüssig)' korrespondiert und weiterführt.
- Auf der Ebene der für den Adressaten zunächst kostenfreien Werbung und Aufklärung wäre die Höhe des Etats sowie die Stückzahl der Mittel ein Indiz. Die Werbemittelstreuung korrespondiert mit dem Grad der (aufgenötigten) Information über die Fachprodukte.
- Auf der Ebene der Beschreibung eines Fachbereichs, der Literatur in einem und über einen Fachbereich (Lehrbücher, Fachzeitschriften, Dissertationen, Fachzeitungen, Fachbücher) spielen Umsatz, Produktstückzahl und -streuung ebenfalls eine große Rolle. Zu unterscheiden sind Produkte für den Insider (Fachmann) und für den Nichtfachmann sowie die fachgezwungene oder freiwillige Käufermotivation.
- Auf der Ebene der Sprache ließe sich die Auswirkung auf die Gemeinsprache und die Verflechtung zwischen beiden ablesen an der Zahl der Wörter, die der Gemeinsprache angetragen, in sie integriert oder in ihr übertragen gebraucht werden. Möglicherweise spielt die Zahl der Wörter mit lexematischer Entsprechung eine Rolle. Möhn spricht in diesem Zusammenhang von "dominanten Fächern".<sup>67</sup>

Diese Kriterien sind exemplarisch und unterschiedlicher Art. Ihr Verhältnis zueinander ist ebenso wenig reflektiert wie die Tragfähigkeit der angegebenen Meßgrößen. Nach den Erfahrungen bei LIMAS verspricht allenfalls eine

Untersuchung der Literatur in einem Fach Aussicht auf Erfolg. Hier können Erhebungen über die Anteile der Sachgebiete an der jeweiligen Jahrestitelproduktion weiterhelfen<sup>68</sup>:

SACHGEBIET	1951	1961	1971	1976	1951 -1976	Gesamt- titelzahl
Allgemeines, Buch und Schrift, Hochschulen	1,5	1,8	2,2	1,4	2,0	14.598
Religion, Theologie	7,4	6,4	4,9	4,6	5,9	43.426
Philosophie, Psychologie	2,2	2,1	2,3	2,7	2,3	17.116
Recht, Verwaltung	5,7	5,8	6,5	5,5	6,5	48.109
Wirtschafts- und Sozialwissenschaften, Statistik	4,5	5,3	8,5	8,0	7,3	54.393
Politik, Wehrwesen	1,6	1,1	2,2	2,0	1,7	12.911
Sprach- und Literaturwissenschaft	2,3	2,9	3,8	3,9	3,6	26.564
Schöne Literatur	17,7	22,9	19,0	21,1	19,1	141.325
Jugendschriften	7,4	4,8	6,3	6,0	5,5	40.436
Erziehung, Unterricht, Jugendpflege	3,1	3,5	4,2	4,9	4,1	30.009
Schulbücher	11,1	9,0	4,0	5,1	6,1	45.074
Bildende Kunst, Kunstgewerbe	2,8	3,0	2,9	3,6	3,3	24.397
Musik, Tanz, Theater, Film, Rundfunk	1,1	1,1	1,6	1,3	1,3	10.039
Geschichte, Kulturgeschichte, Volkskunde	4,0	5,4	4,6	4,0	5,4	40.211
Erd- und Völkerkunde, Reisen	1,7	2,6	2,8	1,9	2,3	17.220
Karten, Kartenwerke	2,0	4,7	4,0	6,1	3,8	28.347
Medizin	3,4	2,3	2,9	3,8	3,2	23.654
Naturwissenschaften	4,4	3,9	5,3	3,9	4,9	36.372
Mathematik	0,5	0,7	1,0	1,6	0,9	7.025
Technik, Industrie, Gewerbe	7,0	4,7	5,4	4,5	5,1	37.735
Verkehr*	3,4	2,5	2,1	0,9	2,1	15.744
Land- und Forstwirtschaft, Hauswirtschaft	2,3	1,3	2,1	2,0	1,9	13.949
Turnen, Sport, Spiele	1,0	0,6	0,9	0,9	0,8	5.916
Verschiedenes	1,0	0,7	0,1	0,1	0,3	2.006
Kalender und Almanache	0,9	0,9	0,4	0,2	0,6	4.514
Insgesamt	100,0	100,0	100,0	100,0	100,0	741.090

\* Bis 1966 Handel, Verkehr

Vergleicht man diese Tabelle mit der oben angeführten Tabelle von Wörterbüchern der verschiedenen Fächer in einer Buchhandlung, so ist bei aller unterschiedlichen Klassifikation doch festzustellen, daß die Fächer, für die Wörterbücher mit mehr als 25.000 Seiten festgestellt wurden, in der Titelproduktion mit Ausnahme der Medizin mit mehr als 5% beteiligt sind. Mögliche Größen zur Berechnung eines Interessenkoeffizienten könnten etwa sein die Anzahl der Wörterbücher des einzelnen Faches, die Anzahl der Stichwörter innerhalb dieser Wörterbücher sowie der Anteil des Faches an der Gesamttitelproduktion und die Verkaufszahlen. Hinzugezogen werden könnte möglicherweise auch die Einschaltquote etwa in den Medien Fernsehen und Rundfunk. Es könnte sein, daß die Anzahl der Bücher einschließlich der Wörterbücher und der Stichwörter in gewisser Relation zu den Textgruppen für das Korpus steht, die das Merkmal öffentliches Interesse haben.

### 2.5.3.3. Interne Schichtung der Fachsprachen

Die mit Hilfe der genannten oder anderer Kriterien ausgewählten Fachsprachen sind weiterhin zu untersuchen auf ihre interne Strukturierung hin. Zu verschiedenen Fachbereichen liegen bereits mehrschichtige Beschreibungsmodelle vor, die als externe Schicht die Verteilersprache, Verbreitungssprache als Sprache des Verkaufs, der Werbung und der Propaganda ansetzen.<sup>69</sup>

Von der Zielgruppe des geplanten Wörterbuchs aus – vom einzelnen Fach aus gesehen der jeweilige Nichtfachmann, der Laie – lassen sich die zu dieser Schicht gehörenden, für das Wörterbuch zentral wichtigen Texte umschreiben mit den Merkmalen: durchgehend thematisch fachorientiert, fachsprachlich relativ stark durchsetzt, für den Nichtfachmann gedacht. Vor allem diese Fachtexte für den Laien sind innerhalb der fachsprachlichen Komponente für das Textkorpus vorzusehen.

Diese enthalten neben Fachsprachlichem auch "Universalsprachliches" und damit Gemeinsprachliches. Der Anteil dieser beiden Komponenten ist weitgehend unbekannt. Doch ergab eine Untersuchung einschlägiger Fachtexte der medizinischen bzw. chemischen Fachsprache, daß beide untersuchten Sprachen Mischsprachen sind, in denen drei Arten von Begriffen unterschieden werden können: Spezifisch fachsprachliche Begriffe oder Terme – Mischterme, deren Bedeutung in der Umgangssprache geläufig ist, die aber einen spezifizierten Gebrauch in der Fachsprache erfahren – umgangssprachliche Begriffe in üblicher Bedeutung. Eine Aufstellung der Texte ergab das Verhältnis 18 : 23 : 125 (Medizin) und 82 : 56 : 182 (Chemie).<sup>70</sup>

Ein weiterer Hinweis ergibt sich daraus, daß in den Häufigkeitslisten der Fachwortschätze Chemie, Mathematik, Medizin und Physik der erste fachsprachliche Ausdruck erst auf Platz 18 (*molecule*), 12 (*function*), 19 (*patient*) bzw. 18 (*energy*) zu finden ist.<sup>71</sup>

Das Verhältnis zwischen vorhandenen und praktisch gebrauchten fachsprachlichen Begriffen wird von Lippert im Bereich der Medizin mit 500.000 : 5.000 angegeben.<sup>72</sup>

In diesem Zusammenhang wäre zu prüfen, inwieweit und in welcher Modifizierung ein Schichtenmodell zur Beschreibung aller ausgewählten Fachsprachen verwendet werden kann, wie die fachexterne Komponente, die Verbreitungssprache in Texten realisiert ist und wie hoch der Anteil des Gemein- und Fachsprachlichen in Fachtexten für den Laien ist. Erst aufgrund solcher Überlegungen ist eine Revision der oben mehr statistisch durchgeführten quantitativen Verteilung von Korpussteilmengen auf einzelne Fächer durchzuführen.

#### 2.5.3.4. Kleinkorpora

In der Diskussion des Verhältnisses zwischen Fachsprachen und Gemeinsprache wurde die Meinung vertreten, es seien homogene Kleinkorpora für die verschiedenen Fachsprachen zusammenzustellen, wobei so lange Material gespeichert werden solle, bis die Wörter erfaßt sind, die man ins Wörterbuch aufnehmen will. Hinzuweisen ist darauf, "daß für alle diejenigen fachsprachlichen Lexeme, deren Verwendungsweisen durch ihren Stellenwert in bestimmten Begriffssystemen festgelegt sind, keine Korpusauswertung vonnöten ist"<sup>73</sup>.

Statistische Untersuchungen haben ergeben, "daß die hundert häufigsten Wörter 60% eines Textes, die tausend häufigsten Wörter 86% eines Textes und die viertausend häufigsten Wörter 97,5% eines Textes ausmachen ("decken")". Aufgrund dieser Zahlen sind die Fachwortschätze verschiedener Fachbereiche untersucht worden. Eine Textdeckung von 84 - 89% durch die häufigsten 1.200 lexikalischen Einheiten hat sich u.a. für Texte des Bauwesens, der Chemie, der Mathematik, der Medizin und der Physik ergeben. Insgesamt sind etwa 15 Subsprachen untersucht worden, wobei ein Stichprobenumfang von 35.000 als günstig gelten kann. Die Stichprobe wurde einem Textkorpus von mindestens 5.000 Seiten wissenschaftlicher Literatur entnommen, und zwar 175 Textausschnitte mit einer Länge von je 200 Wörtern in regelmäßigen Intervallen. Bei den Texten handelt es sich vorwiegend um allgemein verwendete Standardwerke und Hochschullehrbücher mit Überblickscharakter (60%) sowie um nicht allzu spezielle Fachzeitschriften der letzten Jahre (40%). Durchgeführt wurden diese Untersuchungen für die russische, englische und französische Sprache.

Unter den ersten 400 Elementen der Häufigkeitsliste finden sich nahezu alle Präpositionen, Konjunktionen und Pronomen. Daneben gibt es Wörter mit einer ziemlich weiten und allgemeinen Bedeutung, darüber hinaus aber lassen sich unter diesen 400 Wörtern etwa im Bereich der Medizin im Eng-

lischen 56 Substantive und 13 Adjektive ziemlich eindeutig als Lexik medizinischer Texte identifizieren.

Der Vergleich zwischen den ersten 1.200 Einheiten einer fachsprachlichen mit einer allgemeinsprachlichen Worthäufigkeitsliste ergab im Russischen eine Identität zwischen 25% und 37%. Eine stärkere Identität (40 - 50%) ergab sich bei dem Vergleich der Worthäufigkeitslisten der Subsprachen:

Wissenschaft A	Wissenschaft B	Identität
Physik	Medizin	46,5%
Physik	Philosophie	46,0%
Physik	Chemie	50,0%
Philosophie	Medizin	40,08%
Philosophie	Pädagogik	51,7%

Die vorstehenden Ergebnisse dieser Untersuchungen<sup>74</sup> sind u.a. insofern interessant, als die zugrundegelegten Texte nicht spezifisch für Fachleute gedacht sind, sondern eher für Studenten und als solche möglicherweise zu Texten der fachexternen Komponente gerechnet werden können.

Das pädagogische Ziel, dem Fachmann eine wissenschaftlich fundierte Sammlung des Wortschatzes vorzulegen, den er beherrschen muß, um seine Fachliteratur ohne größere Mühe zu lesen, verfolgt auch eine Untersuchung über die Wertigkeit und Frequenz in der Lexis hochspezialisierter medizinischer Texte.<sup>75</sup> Aus zehn maßgeblichen Texten der hochspezialisierten medizinischen Fachsprache werden je 1.000 Wörter von der Seite 34 an berücksichtigt. Aus diesem Korpus von 10.000 Wörtern wird "ein lernwürdiges lexikalisches Minimum von 350 Wörtern" erstellt, das "dem Lehrbuchautor und dem Lehrer in einem Sprachkurs . . . Anhaltspunkte zur Auswahl von Lektionstexten, zur Erstellung von Übungen, Kontrollen und Tests und zur Planung einer lexikalischen Progression bieten" soll. Unberücksichtigt bleiben die unmittelbar verständlichen, verständnisirrelevanten oder vorherrschend strukturbedingten Wörter wie Artikel, Präpositionen, Konjunktionen, Pronomina, Hilfsverben, Modalverben und Pronominaladverbien. Ausgespart werden auch die internationalisierten Maßangaben, Ziffern, Eigennamen und fremdsprachlichen Textteile. Berücksichtigt werden die Verben, Adjektive/Adverbien und die Substantive. In einem Test mit einem Kontrolltext erwies sich, daß die Verbelege zu 51,2, die Adjektiv/Adverb-Belege zu 58,1 und die Substantivbelege zu 43,4% abgedeckt wurden.

Die Untersuchung schließt mit der Empfehlung, "sich von der Registerviel-  
falt eines Mammutcorpus (zu) lösen und sich der Untersuchung von Spe-  
zialregistern mit einheitlichem Mitteilungsrahmen zu(zu)wenden . . . Der  
Arbeitsaufwand wäre geringer, es träten laufend verwertbare und leicht zu  
veröffentlichende Teilergebnisse zutage, und die Spezialminima ließen sich  
im Handumdrehen zu übergreifenden Registern zusammenfassen."

Unter Zugrundelegung von Kleinkorpora für die einzelnen Fächer von  
35.000 bzw. 10.000 Textwörtern ergäbe sich bei einem Gesamtumfang  
der fachsprachlichen Komponente im Korpus von 30 Millionen die Zahl  
von 857 bzw. 3.000 Fächern, die für dieses Wörterbuch berücksichtigt wer-  
den könnten.

#### 2.5.3.5. Das Korpus im Schnittpunkt verschiedener Gesichtspunkte

In den vorstehenden Kapiteln sind einzelne Gesichtspunkte zusammenge-  
stellt worden, die für die Zusammenstellung des Korpus wichtig sind. Ein  
weiterer Schritt wäre, diese Gesichtspunkte zu vervollständigen und die  
Konsequenzen zu überdenken, die sich aus ihrer Kombination für die end-  
gültige Zusammenstellung des Korpus ergeben.

#### 2.5.4. Belege

Unter 2.3.3. ist auf die Komplexität der Belege hingewiesen worden.

Die für sehr oft belegte Wörter genannte Höchstzahl von 2.000 Belegen als  
Grundlage für die lexikographische Bearbeitung könnte "mit Hilfe eines  
Zufallgenerators (als) . . . bearbeitbare Menge von Material aus dem Ge-  
samt" ausgewählt werden.<sup>76</sup>

Zu überlegen wäre, ob man Lemmata, die nur einmal im Gesamtkorpus be-  
legt sind, überhaupt aufführt. Beim Trésor hat man zunächst diese Wörter  
mit aufgenommen, später aber ausgespart.

Wieviele Belege man dann konkret in das Wörterbuch aufnimmt, hängt si-  
cherlich von der Art des jeweiligen Stichworts ab. So ist es sicher so, daß  
Belege bei dem Wort *Tisch* weniger zahlreich sein müssen als etwa bei dem  
Stichwort *Freiheit*. Vom Goethe-Wörterbuch ist bekannt, daß dem Stich-  
wort *Augenblick* für die Bearbeitung über 3.000 Belege zugrunde lagen, von  
denen im Artikel selber 180 Belege zitiert wurden.<sup>77</sup>

#### 2.6. Nachweis der Belege und Verzeichnis der Quellen

Diese Forderung von Grimm läßt sich maschinell insofern leicht erfüllen,  
als die einzelnen eingespeicherten Texte – etwa durch ein komplexes Sig-  
le – hinsichtlich der Quellenangabe gekennzeichnet werden. In diesem Sig-  
le können auch Kennzeichnungen temporaler, arealer, fachbereichlicher

und textsortenspezifischer Art eingespeichert werden. Die Liste der verwendeten Texte läßt sich ebenfalls leicht maschinell erstellen.

Damit ist zunächst die Möglichkeit gegeben, den einzelnen Beleg in einem größeren Kontext nachzuschlagen und bezüglich seiner Beschreibung im Wörterbuch zu kontrollieren.

Darüber hinaus aber gestattet es die maschinelle Speicherung, statistische Erhebungen zu machen, die Bearbeitung bestimmter Lexeme durch die Lexikographen zu überprüfen und Teilmengen mit bestimmten Merkmalskombinationen auszudrucken.

### 3. Ausblick

Grimm hat die Nachweis- bzw. Registrierpflicht begründet mit der Benutzerfreundlichkeit des Wörterbuchs. Die Beachtung dieses Gesichtspunktes ist unter den heutigen technischen Möglichkeiten weitaus umfassender möglich als zu Grimms Zeiten. Darauf deutet die Rede von *äquivalenten Publikationsformen* in These 2 hin – Formen, die nur möglich sind, wenn ein bearbeitetes, maschinelles Textkorpus vorliegt, das unter den verschiedensten Benutzeraspekten ausgewertet werden kann.<sup>78</sup>

### Anmerkungen

- 1 Grimm (1854) I, IV.
- 2 Grimm (1854) I, I.
- 3 Grimm (1854) I, X und IV.
- 4 Bergenholtz / Schaedter (1977) 3.
- 5 Grimm (1854) I, VII.
- 6 Grimm (1854) I, I und IV.
- 7 Weinrich (1976) 367.
- 8 Grimm (1854) I, III.
- 9 Mentrup (1978a) 1.1.1.
- 10 Der Rückgriff der Lexikographen auf Quellen ist – zumindest als Forderung – schon älter (Henne 1968, 93 - 96):
  - 1641 spricht Schottel in Hinblick auf ein vollständiges Wörterbuch des Deutschen davon, "daß ein jedes Wort zu seinem Stamme oder Wurtzel gebracht/ und mit beygefügtten guten Exempeln aus allerhand Autoren erklärt würde".

- 1648 fordert Harsdörffer in seinem Plan eines deutschen Wörterbuchs u.a., daß es vorwiegend auf literarischen Quellen beruhen soll – berücksichtigt werden sollen dabei die Kanzleisprache, die biblische und die literarische Sprache sowie ältere Wörterbücher als Hilfe für die Sammlung der Stammwörter – und daß es die Wörter auch im Kontext, nämlich "redarten, phrasen" bringen soll, "dadurch der gebrauch der wörter erhellet, und müssen solche gezogen werden auß vermeldten Scribenten."
- 11 Grimm (1854) I, III f., X, XVIII, XXIII, XXXV, XXXVIII.
- 12 Grimm (1854) I, XXXIV, XVIII.
- 13 Das *mit* meint, daß noch weitere Grundsätze wie etwa die alphabetische Anordnung, Berücksichtigung auch der "anstößigen" Wörter, die historisch-dia-chronische Ausrichtung für dieses Wörterbuch wesentlich sind.
- 14 Grimm (1854) I, XXXVI.
- 15 Grimm (1854) I, X, I.
- 16 Grimm (1854) I, XXXIV.
- 17 Grimm (1854) I, XXXIV f.
- 18 Grimm (1854) I, XXXV, XXXI f.
- 19 Grimm (1854) I, XXXV. In frühen Briefen Grimms variieren Angaben zur Bandzahl zwischen 6 - 8. Hinweis von A. Kirkness.
- 20 Grimm (1854) I, XXXV, XXXVII.
- 21 Grimm (1854) I, XXXVI f.
- 22 Grimm (1854) I, XXXV.
- 23 Auf dem 1. Colloquium in Bad Homburg (Dezember 1975) wurde in einem Arbeitspapier, IdS (1975), neben vier anderen auch die Einrichtung des Schwerpunktes "Corpora und Belegsammlungen" vorgeschlagen. Auf dem 2. Colloquium (Februar 1976) wurden dann die sechs im Vorwort dieses Bandes genannten Schwerpunkte von dem gesamten Teilnehmerkreis "verabschiedet".
- 24 Für die Berücksichtigung der historischen Dimension bestimmter Stichwörter (vgl. These 19 und 20) ist ein Spezialkorpus bzw. eine durch Exzerption gewonnene Spezialbelegsammlung notwendig, zusammengesetzt etwa aus Geschichtswerken, Schulbüchern, politischen Schriften sowie aus Lexika und Wörterbüchern der entsprechenden Zeitepochen.
- 25 Vgl. im einzelnen dazu Mentrup (1978a).
- 26 E. Weis auf dem 3. Colloquium in Bad Homburg (Mai 1976). Vgl. auch die Thesen 13 und 14.
- 27 Seibicke (1959) 71. Wüster in Drozd / Seibicke (1973) IX. Wüster (1970) 129, 206. Pörksen (1977) 160. Spiegel (o.J.). 1. Zur Medizin vgl. Lippert (1978), Fluck (1976), 91.
- 28 Schewe am 28.5.1976 in Bad Homburg.



- 29 Wüster in Drozd / Seibicke (1973) IX.
- 30 Seibicke (1959) 71. Erben (1965) 18. Drosdowski (1977) 107, 119.
- 31 Wüster in Drozd / Seibicke (1973) IX. Umbach am 21.2.1978 in Hamburg.
- 32 Wahrig (1975). Mackensen (1972). Klappenbach / Steinitz (1961ff.). Duden (1976ff.). Vgl. Drosdowski (1977) 107.  

Zum Vergleich seien zunächst vier einbändige Wörterbücher anderer Sprachen genannt: Webster (1971) erreicht die sehr große Stichwortzahl von 450.000, Robert (1972) enthält 50.000, das russische Wörterbuch von Ožegov (1973) 57.000 und das ungarische von Juhász u.a. (1972) 70.000 Einträge. Vgl. Neubauer (1977) 30f. Zum andern kann man noch hinweisen auf die Zahl der Einträge in Lexika und Enzyklopädien: Für den Brockhaus (1966ff.) in 20 Bänden werden rund 225.000 Artikel angesetzt, für Meyer (1971ff.) in 25 Bänden etwa 250.000 Stichwörter und 100 enzyklopädische Großartikel. Vgl. Schultheiss (1970) 40, 46. Zu den Stichwortzahlen kleinerer und früherer Ausgaben vgl. ebda.
- 33 Buch (1977) 10/11.
- 34 Buch (1977) 19.
- 35 Mentrup (1978a) 1.2.
- 36 Schaeder (1976). Schaeder (1978).
- 37 Erk (1975). Baumbach (1967). Becker (1977). Gebhardt (1976).
- 38 Auskunft von H. Bergenholtz. Gebhardt (1976).
- 39 Karch (1977) 32.
- 40 Der Auszählung liegt ein Korpus von 58 Texten mit ca. 100.000 Stellen der gesprochenen Standardsprache aus dem Korpus der Freiburger Forschungsstelle des Instituts für deutsche Sprache zugrunde. Dies Auswahlkorpus wird von einer Arbeitsgruppe unter Leitung von W. Mentrup auf syntaktische Strukturen hin untersucht.
- 41 Umbach am 21.2.1978 in Hamburg.
- 42 Fachwortschatz Bauwesen (1976) 42. Fachwortschatz Chemie (1973) 47. Fachwortschatz Mathematik (1976) 42. Fachwortschatz Medizin (1973) 45. Fachwortschatz Physik (1976) 45. Vgl. auch Hoffmann (1975) 25. ff. Gebhardt (1976) 283 gibt folgende Rangliste für die ersten zwanzig Wortformen an: *der, die, des, und, in, nicht, den, das, zu, von, im, dem, für, ist, daß, auf, nach, vom, als, eine*.
- 43 Expertise zum Fremdwörterbuch Schulz / Basler von A. Kirkness, E. Link, I. Nortmeyer, G. Strauß.
- 44 Müller (1966).
- 45 Bahr (1962). Bahr (1966). Brief von Bahr am 20.4.1977.
- 46 Umbach am 21.2.1978 in Hamburg. Stickel / Gräfe (1966).
- 47 Dickel am 21.2.1978 in Hamburg.

- 48 Der Aspekt der Quantität scheint maniabler und werbewirksamer zu sein als der der qualitativen Vielfalt: So führen Wörterbuchmacher gegenüber ihren Vorgängern gern die höhere Stichwortzahl ihres Wörterbuchs als Ausweis höherer Qualität an und kommen etwa auf die stolze Zahl von 141.277 gegenüber 55.181 Einträgen, ohne allerdings dabei zu bedenken, daß "die begierde die bei (dem vorgänger) . . . fehlenden . . . wörter nachzutragen" dazu führen kann, "ein wörterbuch anzuschwellen, dasz der gelehrsamkeit entraten konnte, da . . . die 'in jeder minute kreisende und gebärende sprache' dem haschenden, nicht dem stillemsigen samlerfleisz unabhängigen vorschub that", so daß "die hervortretende masse . . . in nichts als zusammen-setzungen (besteht), und wiederum meistens uneigentlichen, denen die art und weise unserer sprache zahllose reihen bilden lässt." Grimm (1854) I, XXIVf, Henne (1972) 56 f.
- So renommierten Wörterbuchverlage gern mit der Anzahl der Stichwörter und der dazu gehörenden Definitionen, Angaben und Belege, wobei dann in der oft flüchtigen Rezeption solcher Daten – vom Verlag unbeabsichtigt? – etwa die um mehr als dreimal so große Zahl der Definitionen zur Zahl der Stichwörter wird.
- 49 Vgl. auch Bergenholtz / Schaefer (1977) 10.
- 50 Weinrich (1976) 363, 365f.
- 51 Weis auf dem 3. Colloquium in Bad Homburg (Mai 1976). Ein ähnliches Verfahren führte in Nancy zur vorläufigen Stichwortliste. Auskunft von I. Vogel im April 1976.
- 52 Vgl. Schaefer (1978), Schaefer (1976). Bergenholtz / Schaefer (1977).
- 53 Bergenholtz / Schaefer (1977) 11.
- 54 Meyer (1971ff.) 867, 868, Stichwort: Buchhandel.
- 55 Gesprochene Sprache (1975).
- 56 Nikitopoulos u.a. (o.J.) 13f. Gesprochene Sprache (1973).
- 57 Die Klassifikation der Wissenschaften als philosophisches Problem, hg. von R. Rochhausen, Berlin 1968, 63. Nach Erk (1972) 28.
- 58 Die Aufzählung der Gebiete mit den Prozentzahlen nach Pfeffer (1975) 16.
- 59 Mittelstaedt / Ströker / Weinrich auf dem 3. Colloquium in Bad Homburg (Mai 1976).
- 60 Buch (1977) 9.
- 61 Erk (1972) 29.
- 62 Dezimalklassifikation (1958).
- 63 Auf dem 1. Colloquium in Bad Homburg (Dezember 1975) von Drosdowski vorgelegt.
- 64 Dahlberg auf dem 3. Colloquium in Bad Homburg (Mai 1976).
- 65 Zur Subklassifizierung des Faches Medizin vgl. Lippert (1978). Hoffmann (1975) 38.

- 66 Bergenholtz auf dem 5. Colloquium in Bad Homburg (Februar 1977).
- 67 Mentrup (1978b) 2.2.4.2. Möhn (1968) 318, 338.
- 68 Gekürzte Tabelle aus Buch (1977) 11 f.
- 69 Z.B. Hoffmann (1976) 262ff. Möhn (1968). Möhn (1976). Mentrup (1976). Mentrup (1978b).
- 70 Rieser: Zur Rekonstruktion der medizinischen Fachsprache. Burghardt: Zur Konstruktion der chemischen Fachsprache. 4. Colloquium in Bad Homburg (Juni 1976). Burghardt (1977).
- 71 Vgl. Anmerkung 42 und die Tabelle oben unter 2.3.3.1.
- 72 Lippert (1978).
- 73 Rosengren a.d. 5. Coll. i. Bd. Hombg. (Feb. 77). Bergenholtz/Schaeder (1977) 10.
- 74 Hoffmann (1975) 25 - 42. Vgl. Anmerkung 42.
- 75 Becker (1977).
- 76 Bergenholtz / Schaeder (1977) 13.
- 77 Umbach am 21.2.1978 in Hamburg.
- 78 Vgl. Mentrup (1978a).

## Literatur

- Bahr, J. (1962): Zur Neubearbeitung des deutschen Wörterbuchs. In: ZFDW 18, 141 - 150.
- — (1966): Technische Verfahren in der Lexikographie. In: ZDS 22, 96 - 111.
- Baumbach, R. (1967): Das Verb in deutschen medizinischen Lehrbüchern. In: Deutschunterricht für Ausländer 17, 11 - 22.
- Becker, N. (1977): Wertigkeit und Frequenz in der Lexis hochspezialisierter medizinischer Texte. In: Zielsprache Deutsch 1, 21 - 27.
- Bergenholtz, H. / Schaeder, B. (1977): Deskriptive Lexikographie. In: ZGL 5.1, 2 - 33.
- Brockhaus (1966ff.): Brockhaus Enzyklopädie in zwanzig Bänden. 17., völlig neu bearbeitete Auflage des Großen Brockhaus. Wiesbaden.
- Buch (1977): Buch und Buchhandel in Zahlen. Ausgabe 1977. Hrsg. vom Börsenverein des Deutschen Buchhandels e.V. Frankfurt am Main.
- Burghardt, W. (1977): Zur Konstruktion der chemischen Fachsprache. Das Lexikon in der Grammatik — Die Grammatik im Lexikon. In: J.S. Petöfi / J. Brede-meier (Hrsgg.): Papiere zur Textlinguistik Bd. 13.1 und 13.2. Hamburg, 135 - 159.
- Dezimalklassifikation (1958): Dreisprachige Kurzausgabe. Berlin/Köln/London.

- Drosdowski, G. (1977): Nachdenken über Wörterbücher: Theorie und Praxis. In: Drosdowski, G. / Henne, H. / Wiegand, H.E. (1977): Nachdenken über Wörterbücher, Mannheim/Wien/Zürich, 103 - 143.
- Droz, L. / Seibicke, W. (1973): Deutsche Fach- und Wissenschaftssprache, Wiesbaden.
- Duden (1976ff.): Das Große Wörterbuch der deutschen Sprache, Hrsg. und bearb. vom Wissenschaftlichen Rat und den Mitarbeitern der Dudenredaktion unter Leitung von G. Drosdowski, Mannheim.
- Erben, J. (1965): Abriß der deutschen Sprache, 8. Auflage, Berlin.
- Erk, H. (1972): Zur Lexik wissenschaftlicher Fachtexte, Verben (= Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts, Bd. 4), München.
- (1975): Zur Lexik wissenschaftlicher Fachtexte, Substantive (= Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts, Bd. 5), München.
- Fachwortschatz Bauwesen (1976): Häufigkeitswörterbuch, Russisch-Englisch-Französisch, Leipzig.
- Fachwortschatz Chemie (1973): Häufigkeitswörterbuch, Russisch-Englisch-Französisch, Leipzig.
- Fachwortschatz Mathematik (1976): Häufigkeitswörterbuch, Russisch-Englisch-Französisch, Leipzig.
- Fachwortschatz Medizin (1973): Häufigkeitswörterbuch, Russisch-Englisch-Französisch, 2., unveränd. Auflage, Leipzig.
- Fachwortschatz Physik (1976): Häufigkeitswörterbuch, Russisch-Englisch-Französisch, 3., unveränd. Auflage, Leipzig.
- Fluck, H.-R. (1976): Fachsprachen (= UTB 483), München.
- Gebhardt, F. (1976): Worthäufigkeiten in der Datenbasis Steuerrecht. In: Datenverarbeitung im Recht 5, 277 - 290.
- Gesprochene Sprache (1973): Berichte der Forschungsstelle Freiburg (= Forschungsberichte des Instituts für deutsche Sprache, Bd. 7), Mannheim.
- (1975): Freiburger Korpus, Textaufstellung, Stand 13. Oktober 1975.
- Grimm, J. (1854): Vorwort in: J. und W. Grimm, Deutsches Wörterbuch, Bd. I. Leipzig, Bd. II. Leipzig.
- Henne, H. (1968): Deutsche Lexikographie und Sprachnorm im 17. und 18. Jahrhundert, In: Wortgeographie und Gesellschaft, Festschrift für L.E. Schmitt, Berlin, 80 - 114.
- (1972): Semantik und Lexikographie (= Studia Linguistica Germanica 7), Berlin/New York.
- Hoffmann, L. (Hrsg.) (1975): Fachsprachen und Sprachstatistik, Beiträge zur angewandten Sprachwissenschaft (= Sammlung Akademie-Verlag 41), Berlin.
- (1976): Kommunikationsmittel Fachsprache, Eine Einführung (= Sammlung Akademie-Verlag 44), Berlin.

- IdS (1975): Arbeitspapier II: Überlegungen zu einem deutschen Wörterbuch, von W. Mentrup unter Mitwirkung weiterer Mitarbeiter des Instituts.
- Juhász, J. u.a. (1972): Magyar Értelmező Kéziszótár, Budapest.
- Karch, D. (1977): Braunschweig – Veltenhof, Zum Sprachinselidiom einer pfälzischen Binneninsel (= University of Nebraska Studies: New Series No. 55).
- Klappenbach, R. / Steinitz, W. (1961ff.): Wörterbuch der deutschen Gegenwortsprache. Berlin.
- Lippert, H. (1978): Fachsprache Medizin. In diesem Band.
- Mackensen, L. (1972): Deutsches Wörterbuch, 7. Auflage, München.
- Mentrup, W. (1976): Gemeinsprache und Fachsprachen. Überlegungen zur Methodik ihrer lexikographischen Erfassung. In: WW 26, 431 - 443.
- (1977): Bericht über den Stand der Überlegungen zu einem großen interdisziplinären Wörterbuchs nach fünf Colloquien. Arbeitspapier, Demnächst als Teil im Forschungsbericht des Instituts für deutsche Sprache "Vorstudien zu einem großen interdisziplinären Wörterbuch der deutschen Sprache des 20. Jahrhunderts".
- (1978a): Überlegungen zur Zusammenstellung und Verwendung eines Korpus für ein großes interdisziplinäres Wörterbuch der deutschen Sprache. In: Bergenholtz, H. / Schaefer, B. (Hrsgg.): Empirische Textwissenschaft, Probleme des Aufbaus und der Auswertung von Text-Corpora. Demnächst.
- (1978b): Überlegungen zur lexikographischen Erfassung der Gemeinsprache und der Fachsprachen, In diesem Band.
- Meyer (1971ff.): Meyers enzyklopädisches Lexikon in 25 Bänden, Hrsg. und bearb. von der Lexikonredaktion des Bibliographischen Instituts. Mannheim/Wien/Zürich.
- Möhn, D. (1968): Fach- und Gemeinsprache. Zur Emanzipation und Isolation der Sprache. In: Wortgeographie und Gesellschaft, Festschrift für L.E. Schmitt, Berlin, 315 - 348.
- (1976): Zur Entwicklung neuer Fachsprachen. In: Deutscher Dokumentartag 1976, 311ff.
- Müller, W. (1966): Die Sprachkartei als Grundlage für die sprachwissenschaftliche und sprachpflegerische Arbeit. In: Die Wissenschaftliche Redaktion 2, 13 - 36.
- Neubauer, F. (1977): Einige Aspekte umgangssprachlicher Lexika. Das Lexikon in der Grammatik — Die Grammatik im Lexikon. In: J.S. Petöfi / J. Bredemeier (Hrsgg.): Papiere zur Textlinguistik Bd. 13, 1 und 13.2, Hamburg, 27 - 64.
- Nikitopoulos, P. u.a. (o.J.): Zur theoretischen Fundierung (von Corpora). Mannheim.
- Ožegov, S.I. (1973): Slovar' Russkogo Jazyka, Moskva.
- Pfeffer, J.A. (1975): Grunddeutsch (= Forschungsberichte des Instituts für deutsche Sprache, Bd. 27). Tübingen.
- Pörksen, U. (1977): Einige Aspekte einer Geschichte der Naturwissenschaftssprachen und ihrer Einflüsse auf die Gemeinsprache. In: Sprachwandel und Sprachgeschichtsschreibung, Jahrbuch 1976 des Instituts für deutsche Sprache (= Sprache der Gegenwart Bd. 41), Düsseldorf, 145 - 166.

- Robert, P. (1972): Dictionnaire alphabétique & analogique de la langue française. Paris.
- Schaefer, B. (1976): Maschinenlesbare Textkorpora des Deutschen und des Englischen. In: DS, Heft 4, 356 - 370.
- (1978): Eine Dokumentation maschinenlesbarer Textkorpora der deutschen Gegenwartssprache. In diesem Band.
- Schultheiss, F. (1970): Bibliographische Anmerkungen zu einer Enzyklopädie und vier Lexika des 19. und 20. Jahrhunderts. In: Die Wissenschaftliche Redaktion 6. Mannheim/Wien/Zürich, 33 - 48.
- Seibicke, W. (1959): Fachsprache und Gemeinsprache. In: Muttersprache, 70 - 84.
- Spiegel, H.-R. (o.J.): Neubenennungen in den technischen Fachsprachen. Manuskript.
- Stickel, G. / Gräfe, M. (1966): Automatische Textzerlegung und Herstellung von Zettelregistern für das Goethe-Wörterbuch. In: Sprache im technischen Zeitalter 19, 247 - 257.
- Tellenbach, E. / Blumrich, M. (1966): Lochkarten als lexikographisches Hilfsmittel. In: ZDS 22, 140 - 145.
- Tollenaere de, F. (1965): Lexikographie mit Hilfe des elektronischen Informationswandlers. In: ZDS 21, 1 - 19.
- Wahrig, G. (1975): Deutsches Wörterbuch. Überarbeitete Neuauflage. Gütersloh.
- Webster (1971): Webster's third international dictionary of the english language. Springfield.
- Weinrich, H. (1976): Die Wahrheit der Wörterbücher. In: Probleme der Lexikologie und Lexikographie. Jahrbuch 1975 des Instituts für deutsche Sprache (= Sprache der Gegenwart Bd. 39). Düsseldorf, 347 - 371. Kurzfassung in: Die Zeit, Nr. 27, 27.6.1975, 34 - 35.
- Wüster, E. (1970): Internationale Sprachnormung in der Technik (= Sprachforum Beiheft 2). 3. Auflage. Bonn.